

How not to respond to *The X-Files*

Belief in pseudo-science is a widespread problem. But some entertainment is misleadingly condemned in that context, promoting the harmful image of the scientist as truth's ultimate custodian.

Science is more than following laboratory protocols. It is about framing hypotheses and devising ways to shoot them down. It is about imagination, of daring to look up from what is known, and ask "What if?" — and then being able to work out the consequences.

That message is all too easily forgotten, sadly, by those who dismiss *The X-Files* as pernicious pseudo-science — and even by those who think more analytically about the television series, at least to judge from an article in the British newspaper *The Independent* (21 August) by John Durant, professor of the public understanding of science at Imperial College, London. As he says, surveys show that the public treats pseudo-science, such as astrology, more as a recreation than a truth to be believed. So why should we be worried about it? Durant thinks that popular fascination with pseudo-science is about the free market of ideas: in which case, he lightly suggests that there should be some kind of consumer organization to tell the public what is truth and what isn't, "checking out the performance of ideas and beliefs".

That notion does the public understanding of science more harm than good. It implicitly casts scientists as heartless, white-coated Gradgrinds with no capacity for imagination — yet scientists know that mental freewheeling is at the heart of their mission. How many rewarding experiments have come from kites flown in the coffee-room?

And now to *The X-Files*. For anyone who has missed it, *The X-Files* is a television drama series (and a movie) about a pair of FBI agents, Mulder and Scully, who investigate paranormal phenomena. Mulder's musings about aliens and suchlike are forever probed by Scully, who shows how the evidence in their casebook could have rational explanations. Very often, the results of their enquiries are inconclusive. The phenomena they observe could have a variety of interpretations, leaving the viewers to make up their own minds.

Is this deliberate looseness an unwitting laxity of plot? On the

contrary, science is more like *The X-Files* than some detractors recognize. It can only progress darkly, up and down many blind alleys and false trails, from hypothesis to hypothesis. If that were not so, science would soon end. Perhaps as important, it invites participation, rather than enforcing the exclusivity of a secular priesthood of which the public would be rightly suspicious.

It is too easy to condemn *The X-Files* as "pseudo-scientific gibberish which would embarrass any self-respecting science-fiction writer" (as Durant polemically does). Again, that misses the point. Science fiction must follow the conventions of any other narrative or dramatic genre: it just so happens that just enough science (not too much) is added to the scenery to make it seem authentic. Yet the interaction of Mulder and Scully is as scientific as you please. They look at the evidence, they come up with hypotheses, they test them, and most are found wanting. Any truth that they think they find is soon undermined by new evidence that their preferred hypothesis cannot explain, and so they are forced to move on.

Why is *The X-Files* so popular? Partly, no doubt, because it responds (or, if you hate it, panders) to the general fascination for the obscure and inexplicable. But, thanks to science, today's oddball notion is tomorrow's orthodoxy. The existence of the inexplicable challenges us to question our hypotheses and, through investigation, extend them. The popularity of *The X-Files* suggests that the public clearly has more of a feeling for the spirit of scientific enquiry than some give it credit for. A rejection of the idea of custodians of truth does not reflect a preference for pseudo-science, but a dislike of being patronized.

The problem that remains is more general: that enjoying pseudo-science is a lot easier than enjoying science, especially if the latter, like the cod liver oil which is unpalatable but Good For You, is administered by a nanny. In the words of Mozart in the film *Amadeus*, who would not rather listen to their hairdresser than to Hercules? □

A chronic lack of definition

Tackling misconduct has been undermined by a lack of resolve of scientific advisers.

It is now six years since the US National Academy of Sciences recommended a new regime of community self-regulation of scientific conduct, five years since the Congress established a Commission on Research Integrity to come up with something better, and three years this November since that commission, chaired by Kenneth Ryan, a former professor of obstetrics at Harvard, issued a report calling for new structures to investigate and adjudicate allegations of scientific fraud.

Ryan's findings found some favour among university administrators, who are currently saddled with what passes for authority over the problem, but were lambasted by scientific leaders. John Dingell (Democrat, Michigan), the *bête noire* of the US scientific community on this and other matters, had lost his bully pulpit at the head of the Commerce Committee in the House of Representatives in 1995, and

the proposals died for lack of support in Washington.

As a correspondent writes on page 823 of this issue, the resultant impasse is undesirable. Yet it is set to continue until such time as a serious incidence of scientific misconduct reignites public concern on the issue. At the rate at which significant cases are arising (see, for example, page 817), that may not take too long.

President Bill Clinton's National Science and Technology Council was asked, in the aftermath of the Ryan report, to prepare a new and universally acceptable definition of scientific misconduct. It has now been sitting on this admittedly difficult problem for two-and-a-half years. It is high time that the council published its agreed definition, and demonstrated that the US government has the courage to tackle this serious problem. □



Arizona professor files lawsuit after being fired for misconduct

[SAN DIEGO] The University of Arizona has for the first time fired a tenured professor for scientific misconduct, spurring both a lawsuit by the scientist and an Internet support campaign.

Marguerite M. B. Kay, who studies Alzheimer's disease and the effect of age on cells, had her contract terminated last month after an investigation found she had manipulated and misrepresented data in three published articles, while mismanaging a laboratory in a way that was "a formula for disaster".

Last week, Kay, 51, filed a lawsuit against the Arizona Board of Regents in the State Superior Court in Tucson to attempt to overturn the university's action.

In an interview, Kay insisted the allegations are "patently untrue", saying she was railroaded by a biased university administrator who sought to oust her at any cost. Arizona officials denied any impropriety, saying academic rules were strictly followed.

Kay said there were some "honest errors" and "a typographical error" in the papers. She said her laboratory staff had used poor research methods without her knowledge. But there was no intent to deceive, she added.

A website — "Witch Hunt at the University of Arizona" — has been set up to solicit



The University of Arizona: faces a court battle after firing a professor for misconduct.

funds for Kay's legal bills. The use of websites is coming into vogue as accused or concerned scientists seek new channels for redress. Kay denied knowing who set up the website, or how much money was received.

As Kay's research was funded by the National Institutes of Health, the federal Office of Research Integrity (ORI) is reviewing the university's action against Kay. After investigating, ORI will issue a report, which could bar Kay from receiving federal funds.

Last May, Arizona's Committee on Academic Freedom and Tenure issued its report

on the allegations against Kay, noting an article in the *Proceedings of the National Academy of Sciences* (PNAS **93**, 5600–5603; 1996) as a prime example of misconduct.

Kay and three co-authors wrote in the article that vitamin E supplements delayed the effect of ageing on the immune system and the brain. This study was the subject of national television news reports.

But the committee's report says the study falsely presented data on mouse immune cells, dropping, selecting or manipulating data points for the desired statistical results.

Insisting this was not true, Kay said there was no evidence to justify the committee's decision.

The report said Kay "hired untrained, mostly undergraduate students with minimal or no prior laboratory experience," and gave little supervision. Kay said people in her lab were correctly trained and supervised.

The lead author of the PNAS paper, Jeffrey E. Poulin, had just received his bachelor's degree in psychology, and two other authors were undergraduates; Kay was the last author. The academy member who sent the article to PNAS was an 87-year-old retired professor, whom Poulin had never met.

Poulin has disavowed the paper's results after learning in the investigation that Kay's research methods were flawed. PNAS editors say they were unaware of the article's deficiencies, which they will now review.

Poulin, 29, explained: "She [Kay] had a big name — like the Michael Jordan of research. I didn't know the way things were [normally] done. She had about 100 publications."

The investigation discovered that lab notebooks had disappeared and computer hard drives had been erased. But Poulin had kept computer copies of the raw data and notebook photocopies, which played a crucial role in the inquiry.

Rex Dalton

Rethink urged on Framework programme

[MUNICH] The European Commission should redesign the proposed research advisory system for its fifth Framework programme, which begins next year, according to the panel that previously advised it on research.

In its last piece of advice, the now defunct European Science and Technology Assembly (ESTA) said the new structure could lead to inefficient coordination of academic and industrial inputs, and confusion about the roles of its committees. In particular, ESTA objects to a plan to split industrialists and academics into separate advisory panels.

In June, the commission said it would set up a structure comprising 17 external advisory groups to oversee the Framework programme's 'key actions', and a two-chamber body to advise on general European Union research policy (see *Nature* **393**, 502; 1998). One chamber will represent academics, the other the industrial community. The chambers will each have 25 members.

Members of ESTA, which comprised both academic and industrial researchers,

were taken aback by the announcement, as they had been appointed only a few months earlier for a three-year period. Nobel laureate Carlo Rubbia resigned in protest.

Last month, other members were further offended to learn that ESTA was to be dissolved immediately, before the new structure has been put in place.

ESTA has given its full backing to the external advisory groups. But it believes the move to create separate chambers for academics and industrialists is retrograde, and will discourage coordination, particularly as the fifth Framework programme is designed to promote economically and socially relevant research.

It is also concerned that all 17 chairs of the external advisory groups will be members of the chambers, believing this could distract discussions from longer term perspectives by introducing more immediate operational concerns.

ESTA is asking for the two chambers to be merged as soon as possible, and for chairs of external advisory groups to be excluded from automatic membership of the combined chamber.

Alison Abbott

Prisoners' DNA database ruled unlawful

[SAN FRANCISCO] A Massachusetts Superior Court judge has barred the state police from demanding DNA samples from prisoners, parolees and probationers. This is one of the first US court decisions asserting privacy rights to stop DNA data banking.

In striking down the statute, Judge Isaac Borenstein said that compulsory blood sampling violated the privacy guaranteed to US citizens under the Fourth Amendment. He impounded 1,200 samples, saying the law's enactment in January amounted to "unreasonable search and seizure".

All 50 US states have laws establishing a forensic DNA database for convicts. Few have been contested in court, and all those have withstood the challenge. Paul Billings, a medical geneticist who supported the Massachusetts plaintiffs, said the ruling showed a balancing trend. "There are beginning to be decisions that show there have to be rules — that all this ethics stuff isn't blowing smoke."

Borenstein ruled that "regardless of the state's compelling interest, an unjustified random bodily intrusion without any indication of individualized suspicion is unreasonable and intolerable".

The state plans to appeal. Assistant attorney general Elisabeth J. Medvedow says the law enforcement value of the database outweighs the intrusion. The DNA bank would help police find missing people, solve crimes and deter illegal conduct, she says.

The statute allows the permanent storage of DNA samples collected from people con-

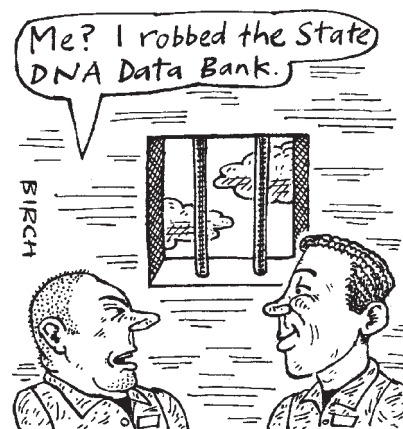
victed of any one of 33 crimes ranging from murder to dissemination of obscenity. The material would be used for law enforcement and — after all identifiers had been removed — for broadly defined research purposes including those that might advance methods of DNA banking and statistical analysis. The plaintiffs argued that this creates a second invasion of privacy, because of the open-ended possibilities for using the data, despite the promise of anonymity.

Frederick Bieber, a forensic geneticist in the department of pathology at Harvard University, testified in favour of the law. He pointed to the high rate of repeat offences in Massachusetts: 26 per cent of convicts who have committed crimes listed in the statute offend again within a year of release.

"From my point of view it's an issue of public safety, victims' rights and preventing future crimes," he says. He adds that permanent storage of samples would be important in the development of improved forensic DNA technology — for example, for use in validating systems using new markers.

State senator James Jajuga, who joined the governor in proposing the statute, says he hopes the database will one day help prevent crimes. Studies of the DNA bank might yield a 'criminal' DNA profile that could help predict which parolees or probationers were likely to commit further crimes, and identify how to use education, drug therapy or counselling as preventive measures, he says.

Jajuga acknowledges the controversy over



such research, but says he feels it could be done fairly. "Obviously we want to be careful with this; there's no question we don't want it to be abused," he says. But Paul Billings questions the science behind hunting for a 'recidivism gene' — and lawyers for the plaintiffs decry the civil-rights implications.

"The law is a blueprint for genetic engineering," says Benjamin Keehn of the Committee for Public Counsel Services in Boston. If prisoners could be tested on the basis of a 26 per cent recidivism rate, he suggests, so could all African American males, who on average have a 28 per cent chance of spending time behind bars.

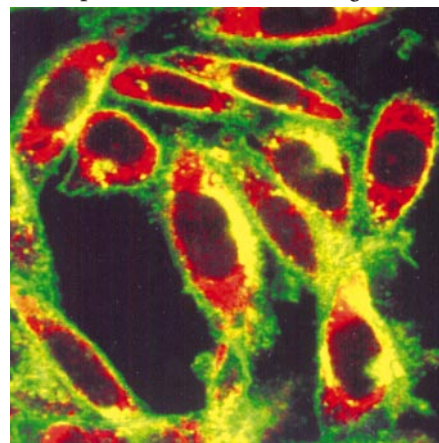
Keehn says that scientists thinking of potentially beneficial results of genetic research need to be aware of how samples are being collected.

Sally Lehrman

NIH institute to work with trial of AIDS vaccine, despite concerns

[WASHINGTON] The US National Institute of Allergy and Infectious Diseases (NIAID) is to collaborate with large-scale, privately funded clinical trials of an anti-HIV vaccine developed by Genentech.

In 1994, NIAID declined to undertake such a 'phase three' trial itself, citing since-



Red alert: trials have begun of a new AIDS vaccine, produced using cells expressing gp120.

resolved concerns about the vaccine's safety — and continuing concerns about its efficacy. The vaccine is based on a part of HIV's protein coat known as gp120 (see *Nature* 369, 593; 1994).

This year, a trial of a modified version called AIDSVAX has gone ahead anyway, led by the Genentech spin-off company VaxGen (see *Nature* 391, 220; 1998). NIAID now says it will collaborate with the trial's sponsors to conduct scientific studies of its own.

NIAID, which is the leading supporter of AIDS research at the National Institutes of Health (NIH), plans, among other things, to bank cells from volunteers. This will allow later analysis of cases of breakthrough infection, and the charting of immune function in successfully vaccinated subjects.

NIAID's new willingness to be associated with the trial has won applause from activists — and barbs from scientists who say there is no evidence that the vaccine will work.

It's "a very positive development," says Jeff Jacobs of the AIDS Action Council in Washington DC. "This alleviates some

concern we had of VaxGen moving forward without the government's involvement."

But some scientists are less sanguine. "What has changed is not scientific, it's political," says Dennis Burton, a molecular biologist at the Scripps Research Institute in La Jolla, California, and an expert on antibody responses to HIV. Another prominent AIDS researcher, who declined to be identified, says: "If anything did come out of the trial, the NIH would be crucified for not having been involved."

Anthony Fauci, the director of NIAID, insists that politics played no part in the decision. The institute will capture scientific information that could answer important questions — such as why the vaccine is ineffective, if that turns out to be the case — and that would otherwise be lost, he says.

"Our interest is based in trying to understand and learn anything we possibly can," says Fauci. "It would be a shame if the only phase three trial [thus far] is executed and we don't get the optimum amount of information."

Meredith Wadman

DuPont opens up access to genetics tool

[WASHINGTON] In the fight over rights to research tools, biomedical researchers received good news last week, as the National Institutes of Health (NIH), the DuPont Pharmaceuticals Company and the Jackson Laboratory reached historic agreements to free up access for researchers to an important tool used in manipulating mouse genes.

The tool, *Cre-loxP*, is a powerful technology that allows target genes to be spliced out of specific cells and tissues. The resulting 'conditional mutants' — as well as several other genetically engineered mice created with the technology — are valuable tools for revealing gene function and, ultimately, for understanding human disease.

DuPont, which holds a patent on the technology, had until last week prohibited researchers who have developed and refined *Cre-lox* mice from sharing them with colleagues, unless the recipients' institutions had signed strict licence agreements. These

included promises to give DuPont exclusive first rights to new *Cre-lox* technologies they developed, and 'reach-through rights' to profits from future discoveries made with the technology.

Dozens of universities had agreed to these terms, but other institutions, including NIH, the Jackson Laboratory in Maine — the largest US breeder of laboratory animals — and the University of California, had dug in, saying DuPont's terms were crushing to the research enterprise. Harold Varmus, the NIH director, led the research community's protest, writing to DuPont last year that it was creating conditions that would "seriously impede" further basic research.

But on 19 August, at a cancer genetics meeting at Cold Spring Harbor Laboratory in New York state, Varmus announced that DuPont and NIH had reached what he described as a milestone agreement in the relationship between academics and industry.

The company has agreed to make the *Cre-lox* technology freely available to NIH researchers and grant recipients for non-commercial purposes. For academic use, NIH researchers will be allowed to disseminate *Cre-lox* materials to colleagues under standard materials transfer agreements. Commercial enterprises will still need to sign commercial licences and pay transfer fees to receive *Cre-lox* technologies developed in the academic community.

Recipient institutions will need to sign agreements with DuPont only if they plan to further transfer the materials or develop them for the benefit of other commercial users. The new agreements will be a far cry from the old ones: DuPont has dropped its claim to reach-through rights on academic discoveries, as well as its rights of first refusal on new technologies.

Furthermore, the dozens of institutions that had earlier signed the more restrictive agreements with DuPont should not worry, as the company has agreed to make the far more liberal terms in its new agreement with NIH available to them.

Paul Friedman, the president of DuPont Pharmaceuticals Research Laboratories, said in a statement that the company's turnaround reflects its commitment to the "wide dissemination of this valuable technology to the academic community".

NIH is hoping that other companies will take the hint. "We hope that the agreement will serve as a prototype for how a commercial organization can put a technology into the academic domain" without losing commercial rights, says Maria Freire, the director of NIH's Office of Technology Transfer.

In a move that is just as important for researchers who have confronted a dearth of sources of *Cre-lox* mice, DuPont has signed an agreement allowing the Jackson Laboratory to collect, breed, preserve and distribute the mice. This is welcome news to scientists such as Jamey Marth, a geneticist at the University of California, San Diego, who helped to develop the *Cre-lox* system. "People who have generated these agents are no longer going to be inundated with requests," says Marth, because the Jackson Laboratory will supply them on a far larger scale.

University officials hope that the agreement will start a sea change in what has been a growing battle over academic access to tools weighted with intellectual property requirements (see *Nature* 393, 505; 1998).

"This changes everything," says Terry Feuerborn, the University of California's executive director of technology transfer. "We've needed something like this to demonstrate that there is room for compromise on the exchange of important research materials."

Meredith Wadman

German transgenic crop trials face attack

[MUNICH] The environment minister of the northern German state of Schleswig-Holstein is leading opposition to two planned field trials of genetically modified crops in the state. The minister, Rainer Steenblock, says he wants changes in federal law to give the 16 *Länder* (states) direct participation in the approval of such trials.

Steenblock's intervention is significant because he is a Green Party minister in a state administration governed by a 'Red-Green' coalition between his party and the Social Democrats. A similar coalition might take control of the federal government in next month's general election, and may then put its weight behind the popular opposition in Europe to the introduction of genetically modified crops.

The field trials, which are being carried out by the German company AgrEvo, involve *Brassica napus*, a species of transgenic rape. They were approved last month by the Robert Koch Institute (RKI) in Berlin, which is responsible for approving such trials throughout Germany.

Steenblock's objections are based on alleged concerns that the transgenes in AgrEvo's rape could be transferred to related crops and weeds, causing ecological problems. He criticizes the RKI for approving the trials without insisting on safety measures, and complains that it adopted a simplified approval procedure for AgrEvo, with no public inquiry, on the grounds that the company is already running similar trials in Lower Saxony.

He cites a case that occurred last spring in the state, in which a genetically modified



Tree trouble: aspen field trial rang alarm bells in Germany over the safety of transgenic crops.

aspen tree began to bud in its third year of growth, instead of its seventh year as is usual. Had the tree bloomed, he says, it would have presented a serious safety problem because the transgene concerned causes stunted growth. RKI's approval of the trial was contingent on the understanding that the trees would not be allowed to cross-pollinate. "This shows how unpredictable genetic engineering can be," says Steenblock.

It is unlikely that the RKI will accept Steenblock's demand for additional safety conditions for trials, however. Ulrich Ehlers, head of the registration office for genetic engineering at the RKI, says the federal authorities were convinced that the experiments were safe.

Even so, AgrEvo is unlikely to avoid a confrontation with Schleswig-Holstein. "If we want to sell these seeds in Schleswig-Holstein, we have to test them there," says a company spokesman. **Quirin Schiermeier**

Ocean drilling project fishes for members

[SYDNEY] The *JOIDES Resolution*, research vessel of the international Ocean Drilling Program (ODP), began its first drilling leg in New Zealand waters this month. But the project — to investigate the circulation of cold Antarctic bottom water — is in urgent need of new members.

Since June, when the Canadian Kathryn Moran took over as ODP director, the programme has relaxed conditions which previously allowed membership only to nations, or consortia of nations, that would pay an annual subscription of US\$3 million.

Speaking in Sydney during the ship's routine change of scientific teams, Moran said that the programme is working hard to cover its costs at the current funding level of US\$45 million. The US National Science Foundation, which pays about half of this, "may pick up the slack for one year but cannot sustain extra funding beyond that," she says.

In 1996, France threatened to withdraw from the programme, questioning its scientific value and perceived US orientation (see *Nature* 379, 193; 1996). Last June, France reduced its contribution by a third and to accommodate this the ODP created an 'associate member' category — which also permitted the admission of China on payment of one-sixth of a subscription.

The United States, Britain, Germany and Japan have renewed full membership and two consortia bring in other members — the European group with 12 partners and the Pacific Rim group which now has four.

When Canada halved its two-thirds share



Kathryn Moran: balancing the books to keep *JOIDES Resolution* afloat after 2003.

in the Pacific Rim group in 1995, the partnership with Australia (one-third) was in difficulty until South Korea joined with one-twelfth and Taiwan contributed one-sixth. Jock Keene, the director of the Australian ODP office, is concerned that the consortium will be unstable as long as it is one-twelfth short of a full subscription. It is being allowed full membership rights as it tries to secure another partner, possibly New Zealand.

Funding cutbacks recently required the Australian Geological Survey Organisation to stop leasing *Rig Seismic*, the only vessel able to conduct marine geology on Australia's continental shelf. Keene says that Australia will struggle to attract more ODP research in Australian waters, because it won't be able to

conduct the surveys needed to prepare for voyages of the *JOIDES Resolution*.

A recent expedition to the Woodlark Basin, east of Papua New Guinea, yielded cores containing bacteria living 842 metres below the ocean floor, almost 100 metres deeper than the previous record. But the voyage highlighted a major deficiency in the ship's capabilities: at an active fault zone on Australia's continental margin, it had to withdraw on detecting signs of hydrocarbons. The vessel is not equipped with pressure-limiting 'risers' to contain the risk of blowout.

Researchers have called for drilling with risers into the hard rock of the deeper crust. Japan is expected to decide this year whether to build a research ship of its own that would be twice the size of *JOIDES Resolution* and able to satisfy this need (see *Nature* 388, 409; 1997). Some see this plan as endangering the ODP after 2003, when the programme and the ship lease are due to be reviewed, but Moran says that a more diverse research programme could be built around *JOIDES Resolution* and the Japanese ship.

Her strategy for now is to seek more fractional members such as India, South Africa, Brazil, Eire and perhaps the US Department of Energy (via its interest in exploring for gas hydrates), and to enter partnerships with consortia of commercial oil companies.

But some ODP participants fear that relaxed membership conditions will encourage other nations to follow France's lead, and try to reduce their subscriptions while retaining a place in the programme. **Peter Pockley**

Spending spree propels Finland towards top of research league

[HELSINKI] Finland is on target to meet its goal of a research funding level of 2.9 per cent of gross national product by next year, as both public and private sources of funds increase at a rate envied by its European neighbours.

The increase in funding places this nation of five million people among the world's biggest spenders on research. The new money is accompanied by policy changes which have led to a greater concentration of effort in areas of strategic importance, such as information technology and molecular genetics, and greater coordination of academic and industrial research.

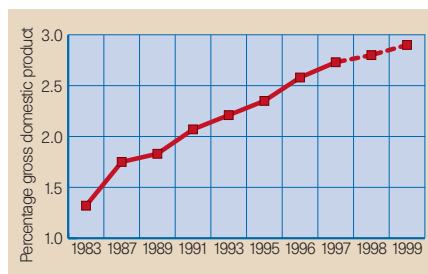
The change in fortune has allowed Finland to implement within 18 months nearly all the recommendations made by the European Molecular Biology Organization which coordinated an external review of Finland's performance in molecular biology (see *Nature* 385, 666; 1997). These include the establishment of a new FM30 million (US\$5.5 million) institute for genomics and bioinformatics research in Helsinki which

coordinates input from three other new university genomics centres, and a similarly organized institute for structural biology.

To support these initiatives the Finnish Academy of Sciences, which distributes much of Finland's basic research funds, has set up new programmes in microbiology, cell biology and transgenics, as well as a five-year programme for genomics, and a smaller programme in microbial ecology. It has made genetics one of the priorities of its new PhD programmes. "The changes were made easy for us, because the new money meant we were spared the difficulties of reallocating money from other areas," says Reijo Vihko, president of the academy.

The ministry for research has recently agreed to maintain direct funding for biotechnology which was due to stop this year.

Finnish scientists seem to be using the new money efficiently. The number of Finnish papers per capita published in refereed journals has doubled over the past decade, overtaking the productivity of



Steady progress: Finnish R&D spend since 1983.

Germany, Norway and the United States. Finland's rate of return from European Union research programmes is the third highest among the 15 states in the union.

In the past few years the academy has managed to lower the average age of gaining a PhD from 38 to 34, and the total number of graduate school positions is set to rise. The experimental centres of excellence programme, which some Finns believe concentrates research funds too heavily for a small country, will be expanded. **Alison Abbott**

Outcry as 'scientific' badger cull is launched to target TB

[LONDON] Conservation and animal welfare groups have launched a campaign to stop a £17.5 million (US\$30 million) government-sponsored scientific experiment that will cull at least 12,500 badgers over five years.

The experiment, announced last week, is designed to investigate the belief, widely held among farmers, that badgers with tuberculosis (TB) pass the disease on to cattle. It will be phased in over the next two years.

The National Federation of Badger Groups (NFBG) says the experiment is "morally unacceptable and practically unworkable". The Edinburgh-based group Advocates for Animals takes a similar view.

Elaine King, conservation officer at the NFBG, claims the experiment could lead to the "total eradication of badgers" in the areas where culling is to take place.

"The government has clearly underestimated the strength of public feeling on this issue," warns King. "While the federation does not condone activities that are unlawful, it is inevitable that members of the public will demonstrate and take direct action."

However, the National Farmers' Union called the experiment "an important step", which will provide hope to farming families "that bovine tuberculosis is on the way to being solved".

Bovine tuberculosis is a relatively small problem in British dairy farming, but it is growing steadily. More than 500 cattle died from tuberculosis last year, compared with 125 in 1991. Infected badgers are widely thought to spread the disease through their use of cattle pastures. However, the suspected link has not been confirmed, and there is no diagnostic test available to detect tuberculosis in living badgers.

The culling experiment is among a package of measures proposed to the government by an expert group of scientists chaired by John Krebs, chief executive of the Natural Environment Research Council.

The government has also agreed to invest in developing a vaccine to protect cattle from tuberculosis, though this will take between 10 and 15 years.

Meanwhile, the experiment will take place in 30 trial sites, each occupying 100 square kilometres in different parts of England where the incidence of cattle tuberculosis has historically been high. Badgers will be lured into traps and then shot. The experiment was designed and will be analysed by a separate group of scientists chaired by John Bourne, professor of animal health at the University of Bristol.

Ten sites will have as many badgers removed as possible. In another 10, only



Losing streak: more than 12,000 badgers will be killed to find out if they spread tuberculosis.

badgers from social groups associated with tuberculosis will be removed. No badgers will be removed from the final 10 sites. On each site, researchers will monitor levels of cattle tuberculosis to see whether, and how much, these are affected by the cull. Laboratory studies will quantify the numbers of dead badgers with tuberculosis.

Jeff Rooker, the government's food safety minister, says the decision to kill badgers was a difficult but necessary one. Bovine TB, he said, needed to be better understood, given the implications for humans who eat meat and drink milk.

But Rooker added that the government had incorporated animal welfare concerns into the experiment. As a result of talks with badger groups, he said, the badgers would be caught in traps instead of snares, which are more efficient but considered to be crueler.

There would also be a closed season between February and April each year to reduce the risk of trapping nursing mothers, ensuring that dependent cubs are not left to starve underground. In addition, farmers would be banned from killing badgers outside the trial areas, which cover less than 0.1 per cent of the area of Great Britain.

But according to the NFBG, these measures do not go far enough. One particular concern, according to King, is the risk of the experiment breaking down through continued — though now illegal — badger culling by farmers. She says many wildlife trusts are likely to refuse culling on their land.

Stephen Harris, professor of environmental sciences at the University of Bristol, who works on badger and bovine tuberculosis, is another critic. He questions the data's end use. Large-scale badger culling would not be a long-term solution even if the experiment confirmed a link between badger tuberculosis and the bovine disease, he claims. And, says Harris, the possibility of tuberculosis passing from cows to badgers also needs to be investigated. **Ehsan Masood**

India's short cow drags Roslin Institute into controversy

[NEW DELHI] The world's smallest cattle breed, nearly extinct in its Indian home, is at the centre of a controversy involving Scotland's Roslin Institute. DNA from the Vechur cow was allegedly smuggled out of India to the Edinburgh institute — where Dolly the sheep was cloned — to be used in patentable research on transgenic animals.

The claims were made by Vandana Shiva, a prominent conservationist and president of the Research Foundation for Science, Technology and Ecology. They were denied by the Roslin Institute. "We have no programme to preserve germplasm from UK or foreign breeds. We don't understand what this is about," says assistant director Harry Griffin.

"I am yet to find out the truth in these allegations," says Moti Lal Madan, director general of the Indian Council of Agricultural Research (ICAR). "Two weeks ago we asked the Roslin Institute for comments and we are awaiting their reply."

The Vechur, no bigger than a large goat, needs little food, thanks to its size — typically 90cm tall and 1m long — but its milk is very rich in fat. Tolerant of heat and resistant to most diseases including foot-and-mouth, it has been the focus of ICAR-funded genetic research in Kerala Agricultural University (KAU) at Thrissur since 1989, when only eight could be found for study. Today there are about 100 on the university farm.

University authorities say that Roslin scientists asked to do research on the cattle in 1994, but that these requests were denied. Shiva alleges that germplasm was later smuggled to the UK by KAU scientists. She also claims that Roslin's web site, a year ago, contained 36 references to its work on the Vechur cow and that these were erased after reports of smuggling appeared in the Indian media.

Dr Sosamma Iype, head of animal genetics and breeding at KAU, says the charges are baseless. She also denies that she spent time at Roslin during 1996, "though our students have been sent there for training," she says. At a press conference this month, KAU vice-chancellor Shyamasundaram Nair denied any wrongdoing by KAU scientists and said that if Roslin had the Vechur germplasm it might have come from other sources.

India's 28 breeds of cattle and 200 million-strong herd are a gold-mine of germplasm, according to Sher Ali, an animal geneticist at the National Institute of Immunology in New Delhi. "We never bother to study their valuable genes until somebody steals them to exploit commercially," he says.

Madan says, however, "If anyone wants a disease-resistant gene it can be found in many other breeds in India." **K.S. Jayaraman**

Salk Institute vet wins \$4.7m damages for wrongful dismissal

[SAN DIEGO] A former veterinarian at the Salk Institute at La Jolla, California, who clashed with scientists there on animal research methods, was awarded \$4.7 million by a jury hearing her wrongful termination lawsuit (see *Nature* 394, 709; 1998).

Teresa J. Sylvina, the institute's veterinarian from 1990 to 1996, should get \$2.7 million in compensatory damages and \$2 million in punitive damages for being maliciously fired, the jury said.

Salk Institute officials, who contended Sylvina was fired for being a poor manager, said the verdict was not supported by the evidence. They intend to appeal. Institute attorneys will also seek to reduce the award. Sylvina, now at Tufts University in Boston, said: "Justice was served."

Call for powerful food safety watchdog in US

[WASHINGTON] The US food safety system should be consolidated under a single powerful authority headed by one individual, according to a report released last week by the Institute of Medicine and the National Research Council.

The report, "Ensuring Safe Food from Production to Consumption", also urges that the federal system be "science-based" and rid of archaic procedures such as the requirement that every chicken and cow carcass be visually inspected.

"Congress should change federal statutes so that inspection, enforcement and research efforts can be based on scientifically supportable assessments of risks to health," the institute says. It adds that the existing, fragmented regulatory structure, involving a dozen agencies, is ill-equipped to meet current challenges in ensuring food safety.

Zoology appointments spark protest in Italy

[MUNICH] A professor of zoology at the University of Pavia has lodged a formal complaint with the Italian research ministry about the results of the recent *concorso* (national competition) for 15 new associate professor positions in zoology.

Many of the posts were allocated to candidates with poor research records in preference to successful researchers, claims Carlo Redi, who is asking research minister Luigi Berlinguer to restart the competition.

Since he took office in 1995, Berlinguer has been trying to introduce more objective criteria into the Italian system of academic promotion, which has been much criticized for favouring personal connections (see

Nature 382, 7; 1996). After two years of political bickering, the Italian senate last month finally approved a watered down form of his new rules. But this came too late for the associate professor *concorsi*.

\$10m grants contest for interdisciplinary research

[WASHINGTON] One hundred academic institutions in the United States have been invited to submit proposals for a new grant programme in interdisciplinary research supported by the California-based David and Lucile Packard Foundation.

A total of \$10 million will be awarded next year, with a maximum single grant of \$1 million. The foundation says it is looking particularly for "novel alliances of disciplines and/or novel approaches to a problem", and will favour projects without funding from other sources. Participation in the programme is by invitation only "to reduce unproductive paperwork" and to increase the chances that a proposal will be funded, according to the foundation.

German universities win more freedom

[MUNICH] Roman Herzog, the German president, last week signed an amendment to the federal framework law on universities, the *Hochschulrahmengesetz*. This will introduce competitive elements to Germany's 325 universities, and give them more financial and administrative autonomy (see *Nature* 388, 820; 1997).

The law comes into effect only four weeks before the federal elections. It will end a two-year controversy over whether the amendment should have outlawed student fees, and whether the changes require the consent of the *Länder* (state) governments. Herzog signed without *Länder* approval.

But this may not be the end of the matter. The Social Democrat Party has announced its intention of proposing a further amendment, including a ban on student fees, if it wins the election. Some *Länder* are considering taking the matter to the constitutional courts to judge whether Herzog was within his rights to sign the amendment without their approval.

Scientists open graves from 1918 flu epidemic

[PARIS] An international team of scientists last weekend opened the Arctic graves of seven miners killed in 1918 by an epidemic of Spanish flu that claimed up to 40 million victims worldwide. The expedition, sponsored by the US National Institutes of Health and led by Kirsty Duncan, a Canadian, hopes to recover virus from the frozen corpses for research.

The team, backed up by the UK exhumation experts Necropolis, has set up extensive safety precautions at the cemetery at Longyearbyen, on the island of Spitzbergen, to prevent the virus escaping. But isolation units have been set up at the local hospital, just in case.

NASA's Big Bang probe on target despite fire

[WASHINGTON] A fire last week at the US space agency NASA's Goddard Space Flight Center in Greenbelt, Maryland, damaged parts of a satellite being built to investigate the faint radiation left over from the Big Bang. But the Microwave Anisotropy Probe, a successor to NASA's 1992 Cosmic Background Explorer satellite, is still expected to be launched in 2000 as planned.

The fire, which started in a compressor, resulted in smoke and water damage to a building in which protective coatings are applied to spacecraft parts. Panels scheduled to be installed on the Hubble Space Telescope in 2000 were also in the building, but were unharmed.

Chinese university buys mini-satellite from UK

[BEIJING] Tsinghua University in Beijing has signed a contract to purchase a £4.8 million (US\$7.8 million) mini-satellite from Surrey University in the United Kingdom. The acquisition is part of Tsinghua's plan to develop itself into a comprehensive research university and to establish a space science programme.

Tsinghua said its choice was based on Surrey's strong reputation for building mini-satellites. The experimental device will be used for communications and observations. Tsinghua will also set up a joint venture with Surrey to manufacture mini-satellites in China. An official signing ceremony will take place in October when Tony Blair, the British prime minister, visits China.

Britain's spies foiled plot to kidnap Niels Bohr

[LONDON] Britain's foreign intelligence agency, MI6, scotched a Soviet plan in 1945 to kidnap the Danish physicist Niels Bohr, according to files recently declassified by the Public Record Office in London.

The files show that intelligence officers obtained details of a plot that involved inviting Bohr to Bornholm, Denmark's easternmost island, where he was to be abducted with the help of Danish communists. Bohr was informed of the plan.

The Danish authorities put a watch on Bohr's house following a Danish intelligence report that uncaptured Nazi war criminals were trying to kill him.

Tougher crackdown on fraud needed

Sir— An increasing number of scientific frauds in US universities are being reported in the media. But do university and federal authorities wish to resolve the problem quickly and effectively? I think not, and I think this is a matter of serious concern.

Fabrication of scientific data, malpractice and violations of federal regulations in university laboratories supported by the National Institutes of Health (NIH) are not only an outrage to all honest scientists, but are also serious federal felonies. Nevertheless, the NIH allows university authorities to carry out an administrative investigation when suspicion of fraud is reported.

University authorities have neither the apparatus nor the authority for a formal investigation: they cannot subpoena witnesses and evidence or seize relevant documents. University committees do not have the legal authority to prevent a defendant's attorneys from resorting to unethical pressure to discourage the whistle-blower and witnesses. Indeed, university authorities themselves are not protected from expensive lawsuits.

It is only at the end of the administrative investigation (which typically takes three to five years) that the university has to inform NIH whether federal grants are involved in the alleged fraud. If the university has decided there has been fraud, it has to pay back all the grants to NIH. This is an unfortunate conflict of interest.

Federal prosecutors inquire into

scientific fraud only if the whistle-blower files a *qui tam*: an action aimed to formally involve the responsibility of federal authorities in the investigation. This usually happens after two to three years, when evidence has long since been contaminated or has disappeared. The federal prosecutor's office is not interested in considering the criminal aspect of the case (violations of federal laws and flagrant fraud in applications for federal grants) or in punishing the defendant. It does not even recommend a probation period during which the defendant cannot receive federal funds. It just collects sufficient evidence to reach a settlement with university attorneys, to recover part of the grants paid by the federal administration.

A recent case at the University of California at San Diego is a classic example (see *Nature* **385**, 566; 1997). In October 1993, the dean's office started an investigation against a professor of medicine for allegations of fabricated research results and violation of federal policies on human and animal experimentation and biosafety standards. In 1995, an *ad hoc* university committee found evidence of fabricated data in at least two articles reporting work supported by NIH grants, but this was not sufficient evidence for misuse of NIH funds: the university biosafety committee sanctioned the scientist for violation of biosafety regulations concerning the use of the AIDS virus. The scientist appealed to the academic senate and in December 1996 a

panel of university scientists (colleagues of the defendant) cleared the scientist of all accusations of scientific fraud or violation of federal regulations.

Interestingly, just one week after the favourable conclusion of the academic senate, the US attorney's office decided to act against the defendant and the university, suggesting that the government was not convinced by the conclusion of the university investigation. About one year later, a settlement of only \$135,000 was reached between the government and the University of California for the reimbursement of federal grants for work containing fabricated data. The defendant was released from any civil and criminal charges. Surely such a low settlement sends a conflicting message to scientists.

Federal legislators must re-examine the problem of scientific fraud — in particular the agreement between NIH and universities — and implement federal judiciary structures to handle from the beginning investigations of scientists accused of fraud (civil as well as criminal aspects), to protect whistle-blowers, and to apply an appropriate punishment in a timely fashion. Such a reform is needed to encourage some scientists to consider more carefully how taxpayers' money is used in their laboratories and to adopt higher standards of integrity.

Paolo del Guercio

1497 Calle Vaquero,
La Jolla, California 92037-7130, USA

Modified animal feeds must be put to the test

Sir— One way to allay public concerns and to find out more about the effect of genetically modified organisms (GMOs) would be to investigate more fully their use in animal feeds. Much money is spent on determining the safety of GMOs as human foods, but would it not be cheaper, easier and more ethical to test animal feeds first?

Large quantities of plant materials, produced by genetic engineering, are destined as raw materials for animal feed: 85 per cent of maize, for example, is used as animal feed or as agro-industrial by-products. Most soya beans are used as protein-rich meal for animals; almost two-thirds of unginned cotton, and most rape seeds and tomato pomace are used as or in feedstuffs. These crops are among the first GMOs submitted for licensing, and will end up in the human food chain. It is obviously

more convenient for research to be done on animal feeds rather than on human food.

The central concept in animal nutrition is 'nutritive value' which is influenced by the presence of undesirable substances, including the potential transfer of harmful factors introduced into the DNA of plants during their conversion into GMOs. Companies base their safety criteria on the principle of 'substantial equivalence' between the engineered and the corresponding conventional plants. To measure this, they generally use chemical, *in vitro* and *in vivo* analyses. Chemical methods compare the sequence of amino acids of the introduced protein with those of known allergens; *in vivo* methods use small laboratory animals for acute oral toxicity tests of relatively short duration. Although these methods are useful tools, one cannot safely extrapolate between species. Biology is often unpredictable: for example the antibiotic cross-resistance to ampicillin in humans. In GMO plants resistant to herbicides, a complex is created

between the 'factor introduced for resistance' and the 'herbicide'. The possibility cannot be ruled out that this complex could be broken down during digestion in the gut or during fermentation, resulting in release of the herbicide.

In addition to the need for labelling and an increased role for legislation and monitoring (guidelines), there is a strong need for research in 'evaluation'. Companies have to demonstrate that GMOs are both effective and non-toxic. Risk assessments are essential to ensure the latter. Study of feeds and farm-animal nutrition for at least one reproductive cycle is also needed. If the health of the animals is not harmed as a result of these tests (which should be done in government-funded institutions), the public is more likely to be reassured. Companies would be in a better position to convince the public of the safety of GMOs.

Pantelis E. Zoiopoulos

Institute of Technology of Agricultural Products,
National Agricultural Research Foundation,
Lykovrissi, Athens 141.23, Greece

White dwarfs sing the blues

Harvey B. Richer

A new model predicts that the oldest white dwarf stars should look blue. That could help to identify what makes up the massive halo of our Galaxy, and to pin down the age of the Universe.

It is curious how events in science often repeat themselves. In 1844, F. W. Bessel noted that the brightest star in the sky, Sirius, appeared to wobble as it moved through space — typical of a star being affected by the gravity of a companion. The companion of Sirius proved to be very faint indeed, but was finally observed by the telescope maker Alvan Clark in 1862, and found to have surprising properties. It was expected to be a cool, red object, as all the low-luminosity stars known at that time were red; but it turned out to be a very blue star, and therefore hot. It was the first 'white dwarf' star to be found. History is now repeated in Brad Hansen's paper on page 860 of this issue¹, where he shows that, contrary to expectations, a certain type of white dwarf star becomes bluer as it ages, not redder. At first sight this may seem rather inconsequential, but it could help us learn about

the contents and age of the Universe.

White dwarfs are the burnt-out cores of middleweight stars. A normal star like the Sun is powered by a sequence of nuclear fusion reactions; once they end, the star ejects its outer envelope to expose its dense core, which cools to become a white dwarf. This white dwarf star will then radiate away its internal heat into space, cooling and becoming less luminous as time passes. But that happens very slowly, and the Universe is too young for even the oldest white dwarfs to have cooled to invisibility.

Because white dwarfs cool at a predictable rate, they can be considered as a kind of cosmic clock: their temperature reveals their age. The difficulty is in how to read this clock. What is the precise relationship between the observed colour of the white dwarf — an indicator of the temperature — and its age? Only theory can provide the

answer to this, so Brad Hansen has developed new evolutionary models of white dwarfs which, for the first time, include a detailed model atmosphere attached to the interior. The atmosphere controls, to a great extent, the spectrum of escaping light, and the rate at which energy leaks out from the star's interior. Hansen finds that hydrogen molecules, which can form in the atmosphere of a cool white dwarf, strongly absorb radiation in the red part of the spectrum, making old, cool white dwarfs appear blue (Fig. 1).

Why should we care whether old white dwarfs are red or blue? Because two fundamental questions in astronomy revolve around the colour and visibility of old white dwarfs: what the Universe is made of, and how old it is.

From the high speeds at which stars are observed to orbit around the centre of our Galaxy, the Milky Way is known to possess a dark halo containing about 90% of its entire mass. The nature of the matter in this dark halo is completely unknown. A measure of the uncertainty is the 70-orders-of-magnitude spread in the masses of the objects proposed to constitute it, from elementary particles up to black holes heavier than 100,000 Suns.

The Microlensing Experiment^{2,3} was designed to resolve this question by monitoring background stars in our nearest neighbour galaxies, the Magellanic Clouds. If a large dark body in the galactic halo passes across our line of sight to one of these stars, its gravity should distort and magnify the background object in a distinctive way (a process called gravitational lensing), telling us the mass of the lensing object. Most astronomers expected either very few lensing events, implying that the dark matter was in the form of particles, or many of them,

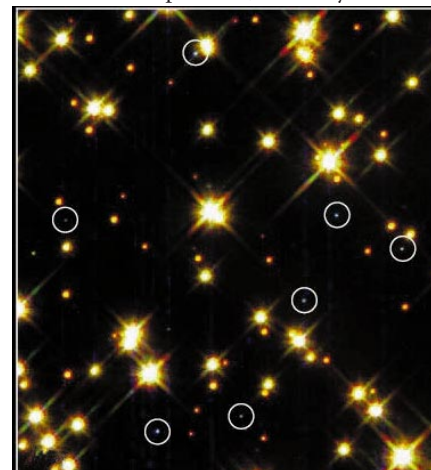


Figure 2 A small part of Messier 4. A few bright white dwarfs are circled, all younger than about one billion years. The oldest white dwarfs in the cluster are about 100 times as faint as these, but it will be possible to use them to put a lower limit on the age of the cluster, and therefore the Universe.

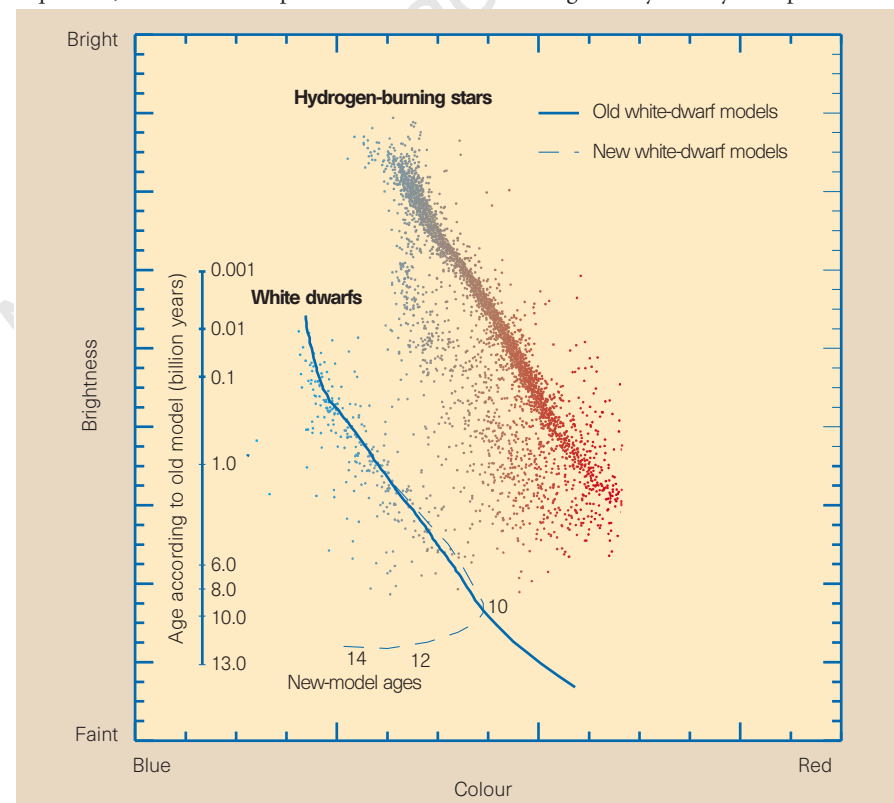


Figure 1 A plot of brightness versus colour for stars in Messier 4, one of the oldest star clusters known in our Galaxy. Old models of cooling white dwarfs, which did not include proper atmospheres, predicted a continuous reddening (solid line). The new models¹ (dashed line) predict that the oldest white dwarfs will be brighter and bluer than expected, making them much easier to find.

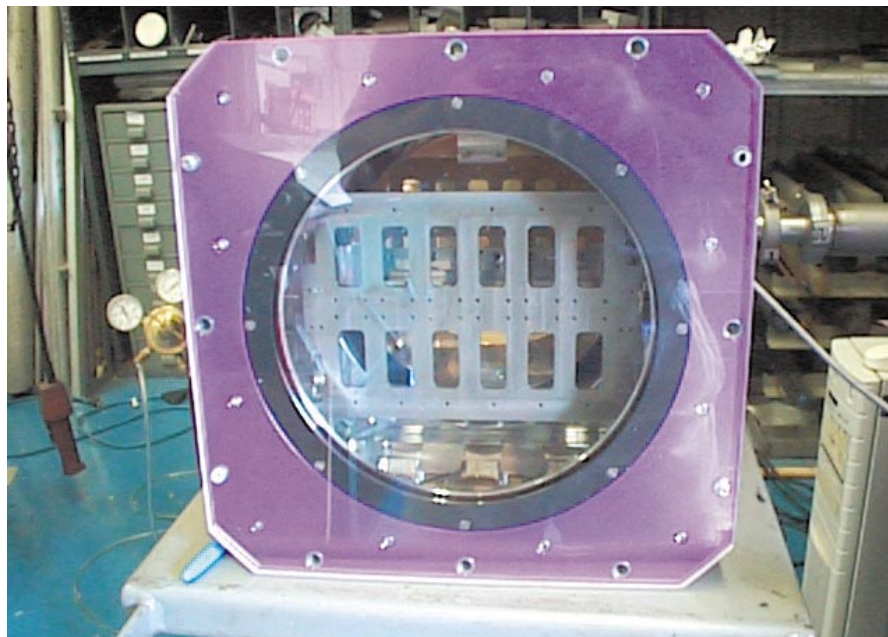


Figure 3 The Canada–France–Hawaii Telescope's large-area camera. This device is carrying out a survey to look for the oldest white dwarfs in our Galaxy, which may form much of the 'dark' matter in the galactic halo. The 12 openings each contain a CCD with eight million pixels. The whole camera covers about 0.3 square degrees of the sky at once, and about 80 different pointings will be made.

implying that the lenses were very low mass stars — perhaps even brown dwarfs, which are too low in mass to ignite nuclear burning in their interiors.

It turns out that there are a lot of lensing events, and the lenses appear to constitute a large fraction of the mass of the Galactic halo — but, surprisingly, the lensing objects are around half the mass of the Sun. If these lensing objects were typical hydrogen-burning stars they would appear as a glowing halo around our Galaxy, which is not observed. The three other possibilities are that these lenses are neutron stars (unlikely, as the heavy metals produced by the neutron stars' precursors would have polluted the Galaxy more than is observed), primordial black holes (also unlikely, but not yet ruled out), or white dwarfs.

If a large fraction of the dark matter resides in white dwarfs it will be possible to observe them. But we must have reliable theoretical models to compare with observations. Current models appear to work fine for white dwarfs younger than about 10 billion years⁴, but our Galaxy is about 12–14 billion years old, so the oldest white dwarfs must be about that age. The new models⁵ bridge this gap.

Successful searches for these old white dwarfs will be made not with the Hubble Space Telescope, as its field of view is too small, but using ground-based telescopes. One large-area survey, now being performed with the Canada–France–Hawaii Telescope (Fig. 3), will cover 25 square degrees of sky to a sensitivity of 25th magnitude. If the entire halo of our Galaxy consists of 12–14 billion year old white dwarfs, all with hydrogen

atmospheres, and if Hansen's new models are basically correct, more than 1,000 old white dwarfs will be found. They will be easily separated from the myriad of old red dwarf stars simply by their colour.

Old white dwarfs can also play an important part in establishing the age of the Universe. One way to determine this age is by measuring the expansion of the Universe; another is to set a lower limit by establishing the age of the oldest stars in our Galaxy. The numbers derived by these independent techniques do not agree. Conventionally

derived expansion ages for the Universe are around 9–10 billion years, whereas the oldest star clusters in the Galaxy are 12–14 billion years old⁵.

Cluster ages are usually found by fitting evolutionary models to the stars that have just completed their hydrogen burning lives. But the physics of these models is quite complex, particularly the nuclear reaction rates, energy transport processes and sources of opacity. A different approach is to search for the least luminous and hence oldest white dwarfs, and calculate the time needed to cool to their present temperature. The physics in these models has its own complexities, but is more tractable than that in models of normal stars. According to the improved calculations of Hansen, these stars will now appear more luminous than we had thought (and therefore easier to find) and their blueness will distinguish them from the huge number of faint, unresolved red galaxies always found on long-exposure optical images.

A real test of Hansen's model will come in the next few years, when early results from the large-area surveys become known, and when images of the nearest old star clusters are taken with the Advanced Camera for Surveys⁶, soon to be installed on the Hubble Space Telescope. □

Harvey B. Richer is in the Department of Physics and Astronomy, University of British Columbia, Vancouver, Canada V6T 1Z4.
e-mail: richer@astro.ubc.ca

1. Hansen, B. M. S. *Nature* **394**, 860–862 (1998).
2. Alcock, C. *et al.* *Astrophys. J.* **486**, 697–726 (1997).
3. Renault, C. *et al.* *Astron. Astrophys.* **342**, L69–L72 (1997).
4. Richer, H. B. *et al.* *Astrophys. J.* **484**, 741–760 (1997).
5. Pont, F., Mayor, M., Turon, C. & VandenBerg, D. A. *Astron. Astrophys.* **329**, 87–100 (1998).
6. Ford, H. C. *et al.* *SPIE* **2807**, 184–196 (1996).

Evolutionary biology

The secrets of faces

Magnus Enquist and Stefano Ghirlanda

The human face is a bewildering source of information, and our abilities to read and remember faces and facial expressions are extraordinarily well developed¹. How faces and our recognition skills have co-evolved, and what information a face can convey, offer many avenues for scientific study.

One thing that faces can tell us is the sex of a person. Although there are only two sexes, we experience many degrees of femininity and masculinity. Perrett and colleagues, reporting on page 884 of this issue², asked human subjects what these differences mean. They manipulated photographs of human faces by enhancing or diminishing differences between the sexes, then they let the subjects rate these manipulated images. The images were produced by first defining a

number of reference points, such as the location of eyebrows, lips, nose and so on. The positions of these points differ in male and female faces, and these differences can be enhanced with computer graphics. For instance, males have bigger jaws than females. So, by increasing the size of the jaws, a more extreme male face is obtained. The authors produced images in which faces varied from feminine to masculine, in this sense, based on the faces of people from Scotland and Japan.

In four experiments, human subjects (men and women from both Japan and Scotland) were asked their opinion about the images with respect to attractiveness, as well as personality traits such as honesty, intelligence and dominance. Female faces were generally judged more attractive when they

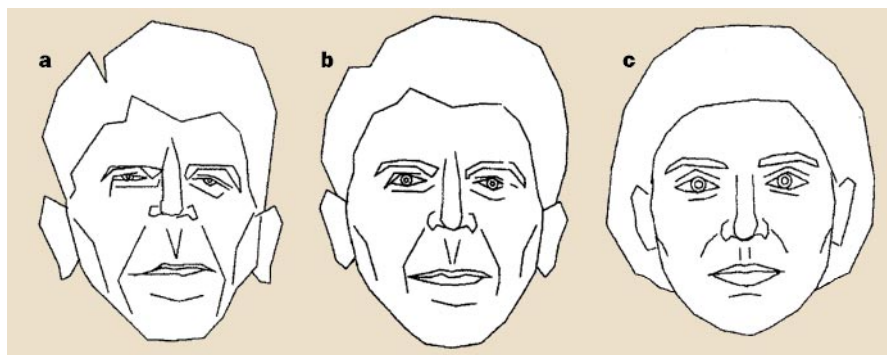


Figure 1 Face to face — three caricatures showing different features. The supernormal caricature of Ronald Reagan (a), produced by Susan Brennan's caricature generator, exaggerates features that are particular to Reagan's face (b), in relation to an average face (c). It is very difficult to give any biological significance to the 'Reagan-ness' of these faces, suggesting that principles of recognition are important in the evolution of signals.

were made more feminine. In contrast, and surprisingly, both men and women preferred slightly feminized male faces over the original or more masculine faces. The Japanese and Scottish subjects more or less agreed in their judgements, with one interesting variation — both groups preferred more exaggeration in faces of their own nationality than in faces from the other country.

When both male and female faces were feminized, subjects rated the person behind the face as more honest, cooperative and emotional. But the result was mixed regarding parental abilities. Whereas the feminized male was rated the better father, the average female was rated a better mother than the feminized female. More masculine faces were thought to be more dominant and older, but judgement of intelligence did not depend on masculine or feminine appearance.

On the basis of these results, Perrett *et al.*² suggest that if a female chooses a male with feminine characteristics, she may get a more honest and cooperative partner who is a better father to her children. The authors also suggest that this might have limited the degree to which male and female faces differ in humans (sexual dimorphism). But masculinity may be an advantage in social competition and dominance — this might explain why male faces do not exactly match female preferences.

Perrett and colleagues have dealt with a classical subject, dating back to Darwin's theory of sexual selection^{3–5}, that is still not satisfactorily resolved. Why do sexual signals look the way they do, and what information do they convey? According to one opinion (favoured in the paper), sexual signals convey important information about the quality of a partner⁴. All aspects of the signal serve this function. This view holds that, during evolution, signals that are reliable cues for fertility, genetic quality (yielding high-quality offspring) and parental abilities have emerged. Males and females have evolved to respond accurately to these cues in partner choice.

A different opinion is that factors related to transmission and recognition are important for the evolution of signals^{5,6}. According to this theory, biases in the sense organs or nervous system influence how we perceive and react to signals. So, sexual signals may just signal sex — the fact that we find some faces more attractive than others may be a by-product of recognition, and may not be linked to partner quality. For instance, it is well known that by altering specific aspects of a familiar stimulus, supernormal effects (stronger reactions)⁷ can usually be produced, even when the new stimulus does not provide the receiver with more information.

As an example of this second view, Fig. 1 shows an average face, a faithful portrait of the former President of the United States,

Ronald Reagan, and a caricature of him⁸. The caricature was produced by an algorithm similar to that used by Perrett and colleagues, exaggerating the differences between the average face and the face of Reagan. This caricature clearly captures and enhances some 'Reagan-ness' from the original portrait. But it is very difficult to ascribe a biological value to such a quality or to argue that we have evolved a Reagan-ness detector.

Thus, before we can distinguish between these two theories, we need to learn more. Studies of faces may provide an excellent testing ground for this. Perhaps the outcome will be a little of both — some aspects of sexual signals will give information about partner quality and others will not. But it is not enough to know what is preferred. We need to find out whether the emotions that faces evoke really do reveal qualities such as parental or social abilities, and we also need to know more about recognition. □

Magnus Enquist and Stefano Ghirlanda are in the Department of Zoology, University of Stockholm, S-106 91 Stockholm, Sweden.

e-mails: magnus@zool.su.se

stefano@zool.su.se

1. Bruce, V. & Young, A. *In the Eye of the Beholder. The Science of Face Perception* (Oxford Univ. Press, 1998).
2. Perrett, D. I. *et al. Nature* **394**, 884–887 (1998).
3. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
4. Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, NJ, 1994).
5. Bradbury, J. W. & Vehrencamp, S. L. *Principles of Animal Communication* (Sinauer, Sunderland, MA, 1998).
6. Enquist, M. & Arak, A. in *Cognitive Ecology* (ed. Dukas, R.) 21–87 (Chicago Univ. Press, 1998).
7. Rhodes, G. *Superportraits* (Psychology Press, Hove, 1996).
8. Brennan, S. E. *Leonardo* **18**, 170–178 (1985).

Earthquakes

A deficit vanished

Steven N. Ward

The southern California earthquake deficit — “Now you see it, now you don’t”, according to an article in the *Bulletin of the Seismological Society of America*¹ by Stein and Hanks. Not done with smoke and mirrors, the vanishing act enlisted a careful revision of our understanding of twentieth-century historical seismicity, and it helped to spirit away a thorny issue that arose in 1995. This was the year that the Working Group on California Earthquake Probabilities² (WGCEP/95) published the report “Seismic Hazards in Southern California: Probable Earthquakes 1994–2024”.

The WGCEP/95 document was remarkable because it struck a new path into earthquake hazard analysis. Previously, geologists and seismologists had independently staked out their own areas of earthquake rate estimation, the heart of hazard calculation. Geologists reckoned the recurrence interval and magnitude of earthquakes by locating

active faults, mapping their length and total offset, and resolving their age. Seismologists concentrated on historical catalogues. Earthquake patterns of the past, they presumed, reflect where and how often earthquakes should strike in the future, and how large they will be. Early geological and seismological studies tended to be piecemeal with few cross-checks. Publication of WGCEP/95 brought order to the field by combining diverse information into a quantitative and consistent multidisciplinary assessment. Space geodesy catalysed the leap forward by providing accurate measures of the pattern and pace of tectonic strain that eventually manifests itself as earthquakes.

Of all the advances in WGCEP/95, one finding seemed to take on a life of its own — the preferred seismicity model predicted twice the number of magnitude 6 to 7 earthquakes than had actually been observed since 1850 (red area, Fig. 1, overleaf). The shortfall

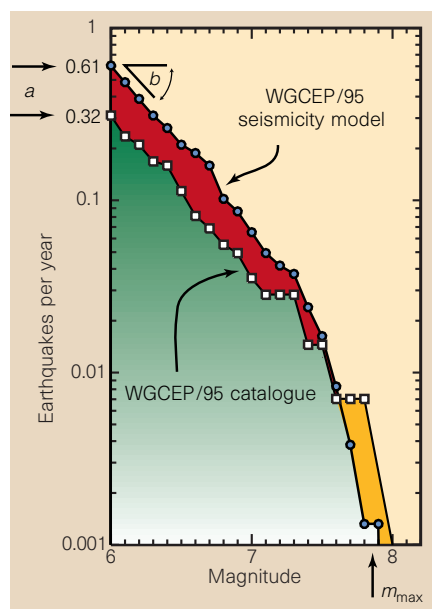


Figure 1 Annual rates of earthquakes greater than magnitude m , for southern California. The lower curve follows the observed rates of earthquakes since 1850 as catalogued by WGCEP/95. The upper curve follows the predictions from the preferred WGCEP/95 seismicity model. The red no-man's-land between the curves is southern California's earthquake deficit, now closed by Stein and Hanks¹, and Field *et al.*^{3,4}. Seismic moment M_T has units of Nm yr^{-1} .

put experts in a pickle. If the earthquake deficit was genuine, then they had to tell California that it faces a far rougher ride than it had experienced in the recent past. If the deficit was an artefact, then one or both of the curves in Fig. 1 was wrong, and those experts had to discover why if they expected to maintain credibility. To address the issue, Stein and Hanks¹ revisited the WGCEP/95 earthquake catalogue. In parallel, Field, Jackson and Dolan^{3,4} dissected the WGCEP/95 seismicity model.

Earthquake rate curves for any region are characterized by four parameters: the a -value, the b -value, the maximum magnitude m_{max} and the total earthquake moment rate M_T . The a -value relates to $n_>(m_{\text{min}})$, the annual number of earthquakes greater than a specified minimum magnitude ($m_{\text{min}} = 6$ for WGCEP/95). The b -value equals the negative slope of $\log[n_>(m)]$ versus magnitude at m_{min} . Magnitude, m_{max} , indicates the largest possible earthquake expected to strike the region. Moment rate, M_T , is proportional to the sum of the slip in every earthquake multiplied by the area faulted. In the long run, the moment rate observed from earthquakes should mirror the product of the region's tectonic strain rate times its area.

Although it would appear to be an objective matter to determine a , b , m_{max} and M_T for each curve in Fig. 1, many factors confound the process. Complicating the observational

side is the fact that the lowest-magnitude earthquakes fix a and b , whereas m_{max} and M_T depend almost entirely on the highest-magnitude ones. Sampling errors cause difficulties at both ends of the magnitude range. Values a and b suffer from incomplete accounting of small earthquakes decades ago when the density of seismic stations was low. Values m_{max} and M_T hinge upon catching the rarest earthquakes. These might recur at intervals longer than California's written history. Multidisciplinary assessments such as WGCEP/95 have an advantage over earthquake catalogues alone in that they employ geodetic strain rates and geological fault data to constrain M_T . Even so, the data inputs and internal modelling decisions that eventually shaped the predicted seismicity curve in WGCEP/95 are subject to considerable uncertainty.

Stein and Hanks¹ conclude that the factor of two mismatch between the rate of observed earthquakes and the rate predicted by the WGCEP/95 seismicity model is not real. The blame splits equally between undercounts of historical earthquakes and overstatements springing from assumptions in the WGCEP/95 model. That is, the lower curve in Fig. 1 should be raised part way, and the upper curve dropped. Stein and Hanks's review of post-1903 earthquakes gives $a = 0.49 \text{ yr}^{-1}$ and $b = 1.0$ compared to $a = 0.32 \text{ yr}^{-1}$ and $b = 0.8$ from WGCEP/95's post-1850 catalogue. They argue that many magnitude 6 to 7 earthquakes were not reported before 1903, which is why they chose the later date as a starting point.

Field *et al.*^{3,4} further identified and removed several elements that inflated M_T in the seismicity model used in WGCEP/95. These included a mistranslation of magnitude into moment and the assumption of

non-Poissonian earthquake recurrence. The most recent seismicity models that draw upon the new analyses^{1,3,4} show no difference between predicted and observed earthquake rates, or between the total moment expended by earthquakes and that inferred from geodesy and geology.

A deficit vanished? One might dismiss this 'news' as much ado about nothing, or conclude that earthquake hazard assessments have yet to evolve beyond magic. In my view, however, these developments demonstrate the new and critical self-evaluation mechanisms available in the current class of multidisciplinary hazard models. In particular, scientists can now recognize where a statistically significant inconsistency exists, quantitatively assess the effects of the various parameters on the inconsistency, and decide whether it is real or artificial. They can then act in response or revise the model, exactly as in the exercises carried out by Stein and Hanks, and Field *et al.*

In truth, of course, unsettled aspects of earthquake recurrence abound even in California, host to the world's most complete earthquake laboratory. Earthquakes will always be difficult to predict because of their elusive and hidden nature. Against these odds, we can make progress only by arming ourselves with the widest range of geophysical weapons. □

Steven N. Ward is in the Institute of Tectonics, University of California, Santa Cruz, California 95064, USA.

e-mail: ward@uplift.ucsc.edu

1. Stein, R. & Hanks, T. *Bull. Seism. Soc. Am.* **88**, 635–652 (1998).
2. Working Group on California Earthquake Probabilities *Bull. Seism. Soc. Am.* **85**, 379–439 (1995).
3. Field, E. H., Jackson, D. D. & Dolan, J. F. *Seismol. Res. Lett.* **69**, 169 (1998).
4. Field, E. H., Jackson, D. D. & Dolan, J. F. *Bull. Seism. Soc. Am.* (submitted).

Translation

Cinderella factors have a ball

Richard J. Jackson

When ribosomes synthesize proteins by translating messenger RNA, they must start the decoding at the correct point on the mRNA. In eukaryotes this is achieved by a scanning mechanism: the small (40S) ribosomal subunit binds to the mRNA near the 5' cap, and then scans in a 5'–3' direction until it encounters the first AUG triplet, which normally serves as the site for initiation of translation¹. This process requires at least nine distinct initiation-factor proteins, composed of a total of no fewer than 25 polypeptide chains.

It is easy to imagine interfering with this complicated machinery by omitting certain factors in the hope of catching the 40S subunit in the actual act of scanning. Imagination is one thing, experiment quite another,

however — hence the appeal of the paper on page 854 of this issue by Pestova *et al.*². They provide the first connections between individual eukaryotic initiation factors (eIFs), notably eIF1 and 1A, and the scanning process.

The factors directly involved in initiation-site selection have hitherto been considered to be eIF2, 3, 4A, 4B and the eIF4F complex. The first step in the process is the formation of a pre-initiation complex: the 40S ribosomal subunit binds eIF3 and a complex of eIF2 associated with the initiator methionyl-transfer RNA (Met-tRNA_i) and GTP (see ref. 3 for further details). This pre-initiation complex then interacts with the mRNA close to the 5' end, probably through the eIF4F complex, which binds to eIF3

NATURE

100 YEARS AGO

It is evident that the moors, valleys and plains of Yorkshire have been depopulated in comparatively recent times. The disappearance of so many conspicuous species is commonly attributed to the glacial period, but I think that the action of man has been still more influential. The extinct animals are such as man hunts for profit or for his own safety. Many of them, among others the cave-bear, *Machairodus*, Irish elk, mammoth and straight-tusked elephant, are known to have lasted into the human period. That so many of them were last seen in the company of man is some proof that he was concerned in their death.

The journey to Tomsk, in Siberia, promises to become quite a pleasant one under the new organisation of the direct trains. The train, which left St. Petersburg on July 31, offered even more comforts to the travellers than the best American trains... It had also, in addition to the usual luxurious fittings of the best Pullman saloon cars, a piano in the first class saloon, a free library provided with a good selection of works on Siberia, as well as with all papers which appear in the towns passed by the train during the journey; a pretty outlook-saloon at the back of the train, with meteorological instruments in it; and even a dark room for amateur photographers, arranged in the second class lavatory. All the furniture is covered with a special material which can be washed with a disinfecting fluid without being injured.

From *Nature* 25 August 1898.

50 YEARS AGO

The curriculum for 'applied medicine', if it can so be called, is always immediately affected by changes in fundamental medical knowledge... Not surprisingly, the medical course has in consequence increased over the past thirty years or so from four to six years. If the same process were to continue, and the strong claims of new interests were admitted, there is no saying how many years would have to be allowed for the training of a doctor in twenty years time.

From *Nature* 28 August 1948.

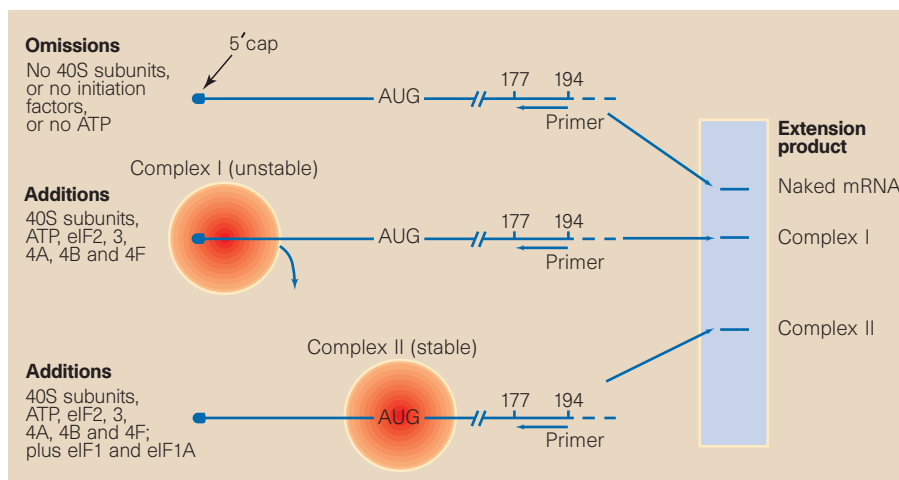


Figure 1 The toeprinting assays used by Pestova *et al.*² to map the position of the bound 40S ribosomal subunit on β -globin messenger RNA. First, initiation complexes were formed by incubation of the components listed on the left; mRNA and Met-tRNA_i were present in all assays, as was a non-hydrolysable analogue of GTP (this stage of initiation needs GTP as a co-factor, but not GTP hydrolysis). Then the primer was added and extended by incubation with reverse transcriptase and radiolabelled dNTPs. The complementary DNA primer extension products were resolved by gel electrophoresis, and the results are shown on the right. The diagram is roughly to scale in terms of the positions of the leading edge of the 40S subunit and of the AUG initiation codon relative to the 5' end.

(itself a component of the pre-initiation complex) and which also has a subunit that binds to the 5' cap.

The 40S ribosomal subunit then scans the mRNA, the scanning being coupled to or immediately following unwinding of the RNA: both eIF4F and 4A have RNA helicase activity dependent on ATP hydrolysis and on initiation factor eIF4B. It is unclear whether the ATP hydrolysis required for initiation is entirely ascribable to the action of these helicases, or whether it is due, in part, to the movement of the 40S subunit itself being driven by ATP hydrolysis. After the AUG initiation codon has been recognized (presumably by pairing with the anticodon of Met-tRNA_i associated with the 40S subunit), the final steps are GTP hydrolysis, the release of all initiation factors associated with the 40S subunit, and the joining of the large (60S) ribosomal subunit. Translation of the mRNA to synthesize the encoded protein then begins.

Why no mention of eIF1 and 1A in this scheme? The reason is that when the initiation pathway was elucidated in the late 1970s, it was found that omission of these two small proteins from the assays caused only a small reduction in initiation efficiency. No well-defined function has been ascribed to either of them and they have often been taken to be non-essential. Nevertheless, the genes encoding both of these factors are necessary for viability in yeast^{4,5}, and so it may be that there was some cross-contamination of the preparations used in the earlier assays.

Pestova *et al.*² have revisited those experiments of the late 1970s, but with three important refinements: they used recombi-

nant eIF1, 1A, 4A and 4B expressed in bacteria; the multi-subunit factors (eIF2, 3 and the eIF4F complex) were from mammalian sources but were almost certainly purer than the earlier preparations; and, instead of studying the interaction of the 40S ribosomal subunit with mRNA solely by sucrose density gradients, the authors used toeprinting assays (Fig. 1) to determine the location of the bound 40S subunit on the mRNA (natural β -globin mRNA).

The striking result was that, when the reaction included ATP and all of those factors thought to be strictly necessary for initiation (eIF2, 3, 4A, 4B and the eIF4F complex), the resulting 40S/mRNA complex (complex I) not only tended to fall apart on sucrose gradient centrifugation, but the toeprinting assay showed that the leading edge of the 40S subunit was positioned just 21–24 nucleotides downstream of the 5' cap (Fig. 1). This implies that the subunit was centred just a few residues from the cap and had not scanned far, if at all. Moreover, complex I must undergo cycles of spontaneous dissociation and (re)formation, because the delayed addition of excess competitor mRNA resulted in the disappearance of complex I formed on the original mRNA.

Most surprisingly, both of the 'Cinderella factors', eIF1 and 1A, were required to produce a complex (complex II), which was not only stable to sucrose gradient centrifugation and to challenge by competitor mRNA, but which had the 40S subunit centred over the authentic initiation codon (Fig. 1). Neither factor alone produced this outcome.

So, is complex I the elusive intermediate in the scanning process? The answer, unfor-

tunately, is probably not. If eIF1 and 1A were added after a short delay, complex I disappeared to be replaced by complex II in similar yield. But delayed addition of excess competitor mRNA together with eIF1 and 1A showed that this was not a case of direct conversion of complex I into complex II on the original mRNA. Rather, the pre-formed complex I must have first dissociated and the appearance of complex II must have resulted from a fresh attempt by the primed 40S subunit to scan the mRNA.

Taken together, these observations provide clues as to the functions — not necessarily mutually exclusive — of eIF1 and 1A. They may be necessary for successful scanning to the initiation codon (but for this they must presumably be present before the pre-initiation complex has interacted with the mRNA); they may prevent the formation of the abortive complex I; or they may accelerate the rate of dissociation of complex I. In the first of these possibilities, it could be that the two factors form an essential part of the 'motor' that drives scanning. An alternative explanation is that they prevent the scanning process pausing at non-AUG triplets, or dissociate any subunits that have stalled at these sites.

Other evidence supports the idea of such a surveillance function, at least for eIF1. First, mutations in the gene encoding yeast eIF1 (originally named *SUI-1*) allowed initiation to occur at a UUG codon. The implication is that wild-type eIF1 normally dissociates the initiation complex, or does not allow it to form, when a mismatched codon/anticodon interaction occurs, and that the mutant form was somehow defective in sur-

veillance^{4,6}. Second, in initiation of translation of encephalomyocarditis virus RNA (which, unusually, is by direct internal ribosome entry at the eleventh AUG from the 5' end, and probably involves no scanning⁷), addition of eIF1 dissociates any complexes formed with the 40S subunit bound at AUG-10; AUG-10 is not used significantly as an initiation site even though it is only eight nucleotides upstream of AUG-11 (refs 2, 7). Finally, and most surprisingly, mutations in yeast eIF1 increase the frequency of misreading of the genetic code within the body of the mRNA (ref. 8), suggesting that wild-type eIF1 may monitor the accuracy of codon/anticodon interaction not only during 40S subunit scanning, but also during the actual decoding of the mRNA sequence.

After the major surprise delivered by Pestova *et al.*, the way forward is not immediately obvious. Undoubtedly, yeast genetics

has a contribution to make, and structural studies of eIF1 and 1A might be instructive. What we would really like to know is which of the other eIFs or ribosomal components eIF1 and 1A interact with during initiation. One thing is certain, however — these two factors will never again be marginalized. □

Richard J. Jackson is in the Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK.
e-mail: rjj@mole.bio.cam.ac.uk

1. Kozak, M. J. *Cell. Biol.* **108**, 229–241 (1989).
2. Pestova, T. V., Borukhov, S. I. & Hellen, C. U. T. *Nature* **394**, 854–859 (1998).
3. Pain, V. M. *Eur. J. Biochem.* **236**, 747–771 (1996).
4. Yoon, H. & Donahue, T. F. *Mol. Cell. Biol.* **12**, 248–260 (1992).
5. Wei, C.-L., MacMillan, S. E. & Hershey, J. W. B. *J. Biol. Chem.* **270**, 5764–5771 (1995).
6. Kasperaitis, M. A. M., Voorma, H. O. & Thomas, A. A. M. *FEBS Lett.* **365**, 47–50 (1995).
7. Kaminski, A., Howell, M. T. & Jackson, R. J. *EMBO J.* **9**, 3753–3759 (1990).
8. Cui, Y. *et al. Mol. Cell. Biol.* **18**, 1506–1516 (1998).

Statistical mechanics

Brownian motion and microscopic chaos

Detlef Dürr and Herbert Spohn

Einstein's 1905 paper¹ on Brownian motion is his most cited article, and one of the most cited physics papers ever. But nearly a century later, we are still learning from this phenomenon. On page 865 of this issue, Gaspard *et al.*² report a high-precision measurement of the position of a colloidal particle suspended in water at thermal equilibrium, providing direct experi-

mental evidence for chaotic dynamics of the fluid on the microscopic scale. Counter-intuitively, perhaps, this strengthens our conviction that chaos underlies the smooth behaviour of matter on macroscopic scales.

At the turn of the last century the atomistic structure of matter, although theoretically well established through the kinetic theory of gases, lacked acceptance because there was little direct experimental evidence for molecular motion. Einstein argued that the erratic path of a small particle suspended in a fluid resulted from the random collisions with the molecules of the fluid. This motion was first systematically described by the botanist Robert Brown in 1828, whence the name Brownian motion. Einstein predicted that its behaviour should be regular, on the average, with the mean square displacement of the Brownian particle growing linearly in time (a hallmark of diffusive behaviour). He related the proportionality factor to Avogadro's number — in other words, to the number of molecules that, according to the atomistic theory, should be around to strike the particle.

Experiments by Perrin in 1909 confirmed this diffusive motion of the Brownian particle, which at the time was regarded as the most direct evidence for the atomistic picture of the world. Kappler repeated the experiments in 1931. He suspended a small mirror in a dilute gas and observed a reflected light signal that looks very similar to Fig. 2 of Gaspard *et al.* on page 866. This experiment led to a determination³ of Avogadro's constant to within 1%.

Immunology

Fetal fascination

A mother readily rejects organ transplants from her offspring. Owing to the contribution of paternal genes, her child's tissues express different antigens to her own — and that's what gets her immune system going. Yet she carries these genetically disparate fetuses to term. What protects a fetus from attack by its mother's immune system? Reporting in *Science* (281, 1191–1193; 1998), Andrew Mellor and colleagues propose that, in mice, rapid consumption of the amino acid tryptophan at the maternal–fetal interface paralyzes the mother's aggressive T cells.

Early in pregnancy, placental cells seem to synthesize indoleamine 2,3-dioxygenase (IDO), an enzyme that degrades tryptophan and can suppress T-cell activity in vitro. Pregnant mice exposed to an IDO inhibitor rapidly aborted their fetuses, unless the embryos were a result of inbreeding and, thus, had a similar genetic make-up to their mothers.



The inhibitor had no effect in mice lacking B and T cells, but when the females were supplemented with T cells their sensitivity to the inhibitor was restored and their pregnancies terminated. So, maternal T cells, directly or indirectly, mediate fetal rejection if they are not kept in check by IDO. Although it is not clear whether exhaustion of the tryptophan supply, or an obscure property of IDO, numbs the maternal T cells, it looks as though mothers do not reject their young because they're IDOalized from the start.

Marie-Thérèse Heemels

All this glosses over a crucial assumption. In his derivation, Einstein simply assumed that successive collisions with the fluid molecules would be statistically independent in approximation. So in essence the position of the Brownian particle is a sum of independent random variables, and as in the theory of random errors that yields diffusive behaviour. But a truly microscopic analysis would have to treat the Brownian particle plus fluid as a dynamical system with very many degrees of freedom, governed by Newton's equations of motion, so deterministic dynamics must somehow account for the assumed statistical independence.

As early as 1918 Smoluchowski⁴ had already identified the instabilities of mechanical motion as the main ingredient for the statistical independence of collisions. Today, an intricate mathematical theory has been developed in which such notions have a precise meaning. It turns out that Brownian statistics, such as the power spectrum found by Gaspard *et al.*, are a generic property⁵ of dynamical systems with so-called hard chaos, in which the motion depends sensitively on initial conditions and there are no traces of quasiperiodic motion.

But, although microscopic chaos is sufficient to produce Brownian motion, it may not be necessary. The analysis of simplified models⁶, such as a hard sphere immersed in an ideal gas and an impurity in a harmonic crystal, has shown that the motion may be Brownian even when the full dynamical system, particle plus fluid, is not chaotic. In these models it is largely the randomness of the initial conditions of the fluid molecules that leads to the erratic motion of the suspended particle.

Gaspard *et al.* show that the fluid in their experiment has chaotic dynamics after all. This points at an apparently general property of systems with many particles. From experience we know that fluids effectively maintain a well-defined local temperature and pressure. Of course, there can be turbulent motion, but even then a small fluid element is approximately in thermal equilibrium, and that is the basis of all of hydrodynamics. Also, external perturbations hardly change such a state of local equilibrium. If the motion of the fluid particles were quasiperiodic (governed by a finite number of frequencies) this stable macroscopic behaviour would be hard to explain.

Only chaotic mechanical motion generates enough intrinsic noise to ensure a robust average behaviour. Chaos is usually defined mathematically in terms of positive Lyapunov exponents, which characterize the exponential divergence of nearby trajectories in phase space (this exponential divergence is what makes these systems sensitive to initial conditions). For many-particle systems, the evidence points to a spectrum of

Lyapunov exponents that scales with system size and has a positive part. Nevertheless, there can be pockets in phase space with quasiperiodic motion, where all Lyapunov exponents are zero, especially when forces are attractive. Presumably, such a mixed phase space will have little effect on the observed macroscopic features of fluids, because observable properties such as local temperature, velocity, and pressure are not sensitive to such fine dynamical details — but a great deal of effort is still required to understand, on the microscopic level, what degree of chaos in a mechanical many-particle system

is needed in order to ensure the regular macroscopic behaviour we see around us. □

Detlef Dürr is in the Department of Mathematics, Universität München, and Herbert Spohn is in the Centre for Mathematics, Technische Universität München, D-80290 München, Germany.

e-mails: duerr@rz.mathematik.uni-muenchen.de
spohn@mathematik.tu-muenchen.de

1. Einstein, A. *Investigations on the Theory of the Brownian Movement* (Dover, London, 1956).
2. Gaspard, P. *et al.* *Nature* **394**, 865–868 (1998).
3. Barnes, R. & Silverman, S. *Rev. Mod. Phys.* **6**, 162–192 (1934).
4. Smoluchowski, M. *Naturwissenschaften* **17**, 253–263 (1918).

Mimicry

Sheep in wolves' clothing

Graeme D. Ruxton

Avoiding being eaten is a vital activity for many animals. They can in part do so by looking like other, unpalatable species, and the study of this strategy (known as mimicry) has a long and distinguished pedigree — so long, in fact, that from the textbooks one might think that the topic was pretty much sewn up. In a paper in *Animal Behaviour*¹, however, MacDougall and Stamp Dawkins come up with a new angle. They argue that the ability of a predator to discriminate between types of prey has to be taken into account. The complications that introduces may well require revision of thinking on mimetic systems.

Aposematic coloration, where unpalatable, poisonous or dangerous animals adopt conspicuous markings, is a well-known phenomenon. The usual explanation is that predators learn to associate these markings with unpleasant experiences, and so are less likely to attack similarly marked individuals in future. This mechanism leads to selection pressure for similarity of markings between species, and hence mimicry; this can be seen in the many insects, some of them completely harmless, which have the yellow and black stripes characteristic of wasps. Traditionally

mimicry is divided into two types: Müllerian, where two unpalatable species both benefit from sharing the mortality costs of predator learning; and Batesian, where a palatable species benefits from its resemblance to an unpalatable species, which in turn pays a cost because the palatable mimic degrades the quality of its aposematic signal.

There are two major objections to this story. First, it has been argued that true Müllerian mimicry can exist only in highly specialized and unrealistic circumstances², because there will always be a difference in palatability between two species and so the less palatable one will always be disadvantaged by the mimetic relationship.

Second, theory predicts that Müllerian mimics should be monomorphic, because the more similar-looking unpalatable individuals there are, the more there are to share the costs of attacks by naive predators. In contrast, the palatable Batesian mimics should be polymorphic; the reason is that the protection individuals of a given morph get from mimicking an unpalatable species decreases with the frequency of palatable mimics. It is strange, then, that several unpalatable (and so, by the definition

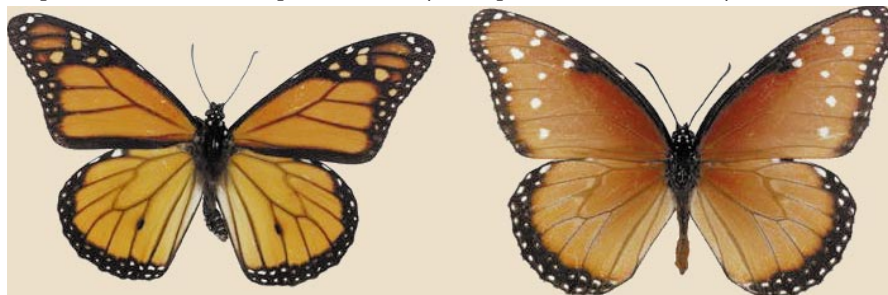


Figure 1 Morphs and mimicry. The viceroy butterfly, *Limenitis archippus*, occurs as two morphs and was once thought to be a polymorphic Batesian mimic. It turns out, however, that it is less palatable to its natural predators than the species that were assumed to be its models — the monarch (*Danaus plexippus*, left) for one morph and the queen (*Danaus gilippus*) for the other. In conventional theory, it is hard to explain cases of polymorphic Müllerian mimicry such as this. But the new ideas of MacDougall and Stamp Dawkins¹ may account for its occurrence.

above, Müllerian) mimics are polymorphic³ (Fig. 1).

The theory proposed by MacDougall and Stamp Dawkins¹ offers a resolution to both problems. The authors accept that the less palatable species will generally be disadvantaged by its resemblance to the more palatable one (what we could term a misidentification cost). But they suggest that there can be a hitherto unconsidered advantage to mimicry which benefits both parties, and which may outweigh the misidentification costs paid by the less palatable species: mimicry reduces the risk to individuals from both species of being misidentified by the predator as belonging to a third and even more palatable prey species (that is, it can provide a misidentification benefit). Predators often have a wide diet, so they have to differentiate between a large number of categories of prey appearance and associate a palatability score with each category. Because of cognitive limitations of the predator, the more categories there are, the more mistakes will be made.

MacDougall and Stamp Dawkins propose that mimicry reduces the number of categories, and therefore may reduce the frequency of mistakes where unpalatable species are misidentified as palatable. Thus, they claim, Müllerian mimicry, where both species benefit from their similarity, is possible in realistic circumstances. Further, the resemblance between two unpalatable species could be either Müllerian or Batesian, depending on the relative strengths of the misidentification costs and benefits, which could explain the existence of unpalatable but polymorphic mimics.

The conceptual leap here is that the mimetic relationship between two species cannot be understood without proper consideration of their shared predator. Whether the relationship is Müllerian or Batesian will depend on the discrimination abilities of the predator, and also on the properties of alternative prey. The relationship could be different for different predators, or for the same predator at different times or in different places.

All this is theory, however, and experiments on learning and cognition in predatory species will be needed⁴. This work is likely to be best achieved initially in the laboratory, but several predictions are ready for testing in the field. For example, the new theory has it that polymorphic but unpalatable mimics do not benefit strongly from mimicry that produces a reduction in predator misclassification errors. This could arise because the common predators of these species feed on only a few species or have highly developed discriminatory powers. Field investigation of the foraging behaviour of these predators could be fruitful.

There is also scope for taking theory further. Misidentification is costly to the predator, so we should not, as in the past, consider

the predator's abilities as fixed and then ask how these abilities influence the evolution and maintenance of mimicry in prey species. Rather we should ask how prey mimicry and predator discrimination co-evolve and co-exist. Because it now appears that the operation of both Müllerian and Batesian mimicry require mistakes by the predator, the results of this enquiry should be both subtle and illuminating. □

Synaptic plasticity

Down with novelty

Richard G. M. Morris

"Too many new ideas around here," I tease my colleagues, when unfinished but much-loved experiments are dropped in favour of pursuing something else dreamed up the night before. "But you always say that the best experiments start in the pub," they intone, "...and it's different from the same old stuff we've been doing for years." Duly defeated, I frown my way back to the office wondering whether the new is always destined to drive out the old. An intriguing observation on synaptic plasticity, reported by Xu *et al.*¹ on page 891 of this issue, suggests that this old aphorism may indeed have some physiological basis. They have investigated what happens when animals in which long-term potentiation (LTP) has recently been induced are transferred from a familiar environment, where electrical recordings from the brain have hitherto been made, into a new one. And they find that LTP disappears rapidly as animals explore a new place and, presumably, encode new information.

LTP is widely held as a model of the activity-dependent synaptic plasticity that may underlie the automatic encoding of new information^{2,3}. One area of interest is the mechanisms of plasticity that are responsible for decreasing the efficacy of synaptic transmission. Such decreases could occur immediately after LTP is induced, in a phenomenon called depotentiation. Or, they could occur from a baseline at which synaptic efficacy has already stabilized, as in long-term depression.

Attempts have been made to identify patterns and frequencies of stimulation that lead to these decreases. Long trains of low-frequency stimulation⁴ or pairs of pulses⁵ are effective *in vitro* and sometimes⁶ (although not always) *in vivo*^{7,8}. Stimulation at 5 Hz is particularly effective in triggering depotentiation within a short time after induction of LTP, and this may involve the destabilization of synaptic interactions mediated by integrin receptors⁹.

Xu *et al.*¹ investigated whether behavioural manipulations, rather than artificial

Graeme D. Ruxton is in the Division of Environmental and Evolutionary Biology, University of Glasgow, Glasgow G12 8QQ, UK. e-mail: g.ruxton@bio.gla.ac.uk

1. MacDougall, A. & Stamp Dawkins, M. *Anim. Behav.* **55**, 1281–1288 (1998).
2. Speed, M. P. *Anim. Behav.* **45**, 571–580 (1993).
3. Owen, D. F., Smith, D. A. S., Gordon, I. J. & Owiny, A. M. *J. Zool. Lond.* **232**, 83–108 (1994).
4. Turner, J. R. G. & Speed, M. P. *Phil. Trans. R. Soc. Lond.* **B351**, 1157–1170 (1996).

trains of electrical stimulation, could have similar effects. They first recorded the brain activity of rats in a small — but well-lit — Perspex box with which the animals were familiar. The rats were then transferred to a box with different lighting conditions and fresh wood shavings on its floor. An important two-pathway control established (although it is unclear how many of these experiments were conducted) that exposure to a new environment could depotentiate recently induced LTP, without disturbing an independent but non-potentiated pathway.

These results rule out the potential contribution of artefacts (such as changes in brain temperature of the exploring rats), and reveal that the underlying mechanism shows input specificity. Merely handling the animal shortly after induction of LTP did not cause depotentiation, suggesting that it may be necessary for the rat to explore the new place. However, exploration alone is insufficient, because synapses at which LTP had been induced the day before could not be depotentiated, even though the rats explored the new place just as much as the old one. Stress is also unlikely to be important, because the new environment was deliberately made non-stressful, and the rats showed neither behavioural nor hormonal indices of stress. The emerging picture is that non-stressful exploration of a new place can cause recent synaptic potentiation to be reset, without affecting synapses at which LTP has already stabilized.

A puzzle about Xu and colleagues' findings is the apparently immediate effect of exposure to the new environment. Given the relatively long trains of stimulation that are needed to induce depotentiation in physiological studies, one might have expected a more gradual time course to exploration-induced depotentiation than the apparently sudden decrease to near baseline within a few minutes. It is also slightly puzzling that the peak of the electroencephalographic activity at 6–8-Hz that accompanied exploratory movements did not seem to occur until after the depotentiation had already been triggered.

A critical experiment that could now be done is to exploit the 'down with novelty' phenomenon, to explore the putative relationship between LTP and memory. Such a study would examine whether exposure to novelty accelerates the rate at which recently acquired information is forgotten, and what types of novelty are effective in doing this^{10,11}. Any one-trial, hippocampus-dependent memory task would be suitable. However, the difficulty with any simple relationship between novelty and hippocampal memory function is that certain natural behaviours, such as food caching in scatter hoarding mammals, involve many items of new information, all of which have to be remembered for a time¹². A memory system that could recall only the last item cached, wiping out memory of earlier items, would be unhelpful.

Because humans and animals can readily recognize and remember many novel items in laboratory situations, event novelty and context novelty may have different effects. Perhaps many new items can be successfully encoded and recalled within a single familiar environment, whereas exposure to a novel environment shortly thereafter will threaten successful recall of items recently stored in the familiar environment. Even so, novelty is unlikely to be the sole determinant

of hippocampus-mediated encoding or forgetting — returning to the example of food caching, cache retrieval resets memories. Animals do not waste time returning to places from which caches have already been retrieved.

For the present, the electrophysiological data indicate that the new can drive out the new, but not necessarily the old. Xu and colleagues' findings are another novel twist on the yellow-brick trail of the engram. □

Richard G. M. Morris is in the Department of Neuroscience, University of Edinburgh, Crichton Street, Edinburgh EH8 9LE, UK.
e-mail: r.g.m.morris@ed.ac.uk

1. Xu, L., Anwyl, R. & Rowan, M. J. *Nature* **394**, 891–894 (1998).
2. Bliss, T. V. P. & Collingridge, G. L. *Nature* **361**, 31–39 (1993).
3. Morris, R. G. M. & Frey, U. *Phil. Trans. R. Soc.* **352**, 1489–1503 (1997).
4. Dudek, S. M. & Bear, M. F. *Proc. Natl Acad. Sci. USA* **89**, 4363–4367 (1992).
5. Thiels, E., Barrionuevo, G. & Berger, T. J. *Neurophysiol.* **72**, 3009–3116 (1994).
6. Heynen, A. J., Abraham, W. C. & Bear, M. F. *Nature* **381**, 163–166 (1996).
7. Errington, M. L. et al. *J. Neurophysiol.* **74**, 1793–1799 (1995).
8. Doyere, V. et al. *Hippocampus* **6**, 52–57 (1996).
9. Staubli, U., Chun, D. & Lynch, G. J. *Neurosci.* **18**, 3460–3469 (1998).
10. Knight, R. T. *Nature* **383**, 256–259 (1996).
11. Honey, R. C., Watt, A. & Good, M. J. *Neurosci.* **18**, 2226–2230 (1998).
12. Jacobs, L. F. in *Behavioural Brain Research in Naturalistic and Semi-Naturalistic Settings*. (ed. Alleva, E.) 301–322 (Kluwer Academic, Amsterdam, 1995).

Evolutionary ecology

Bedazzled by flowers

Nick Waser and Lars Chittka

Human eyes are dazzled by the diversity of flower colours, and early biologists naturally interpreted colour as a signal for other eyes — those of pollinating insects. A close look at flowers reveals surprising additional detail. Petals of the snapdragon *Antirrhinum majus*, for example, have a remarkable sparkling sheen caused by epidermal cells with a conical surface. In a recent issue of *Heredity*, Glover and Martin¹ argue that this sparkle adds to background flower colour in bedazzling and attracting bees, and their study is a welcome addition to a surprisingly small number of experiments that have explored the great diversity of floral features in higher plants.

Glover and Martin studied individual and compound effects of petal sparkle and overall colour. They used mutants at the *mixta* gene locus, which cause epidermal cells to develop a flat surface, and the *nivea* locus, which confer white rather than the normal magenta petals. The authors planted wild type, single mutant and double mutant snapdragons in a garden exposed to bees, and discovered that both *mixta* mutants (with dull petals) and *nivea* mutants (with white petals) suffer lower reproductive fitness, measured as the pro-

portion of flowers that produce fruit.

Glover and Martin conclude that bees prefer to visit sparkling, magenta flowers, thus increasing the fruit set. An alternative that would be useful to explore is that *mixta* or *nivea* genes have pleiotropic effects on characteristics other than colour, so that natural selection favours sparkle and bright colour indirectly, through the fitness value of these correlated characteristics. For example, the biochemical pathways leading to the production of flower pigments also generate

other end products that may affect survival, growth and defence^{2,3}. Reports of such correlated selection are increasingly common, and include examples of flower colour^{2,3} and other floral features⁴.

If selection does act directly through pollination, this could take several forms. Mutants might attract a different mixture of bee species than wild-type flowers. Or, they might elicit fewer visits or other behavioural changes in the same species of bee. Although direct observations of snapdragon pollination remain to be done, advances in cognitive neurobiology provide some fascinating suggestions as to how bees might respond to the different genotypes.

A preference against the white *nivea* mutants, which reflect ultraviolet light¹, is easily predicted — such flowers resemble green foliage to a bee's eye (both are 'bee-uncoloured'; see Fig. 1). A distinction that our eyes make easily is difficult for bees, whose colour perception does not code brightness⁵. A preference against *mixta* mutants could be predicted on different grounds. These mutants appear pink compared with the magenta ('bee-blue') wild-type petals, and they also lack sparkle. The angular resolution of a bee's eye is far too coarse for it to see sparkle⁶, but bees easily learn to distinguish petals of different textures with their antennae or feet⁷. The bees lack any innate bias in this tactile sensory modality, but they learn to associate texture with superior nectar or pollen reward in flowers.

Bees can also distinguish pink ('bee-bluegreen') *mixta* mutant flowers from wild-type flowers by means of their colour vision (Fig. 2, overleaf). Should *mixta* flowers offer a poor reward as a pleiotropic effect of the mutation, bees would learn to discriminate against them. Also, there is some evidence for an initial bias by naive bees against bee-bluegreen flowers, and this might have evolved because plants with such flowers often provide an inferior floral reward⁸. But this bias is easily overcome by learning⁷.

These possibilities contrast with a classical assumption that different flower colours

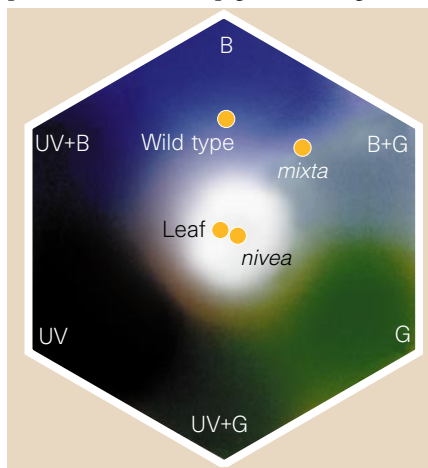


Figure 1 How bees perceive colour. The point generated by a coloured object in the colour hexagon informs us how bees will perceive the object through their ultraviolet (UV), blue (B) and green (G) receptors, and through further processing of receptor signals in the central nervous system. Wild-type *A. majus* flowers are 'bee-blue', whereas *nivea* mutants are perceived as 'bee-uncoloured'. These mutants are hard to detect against green leaves, which fall into the same colour category for bees. 'Bee-bluegreen' flowers, such as the *mixta* mutants, are discriminated against, either because of an innate bias or because they offer less reward.

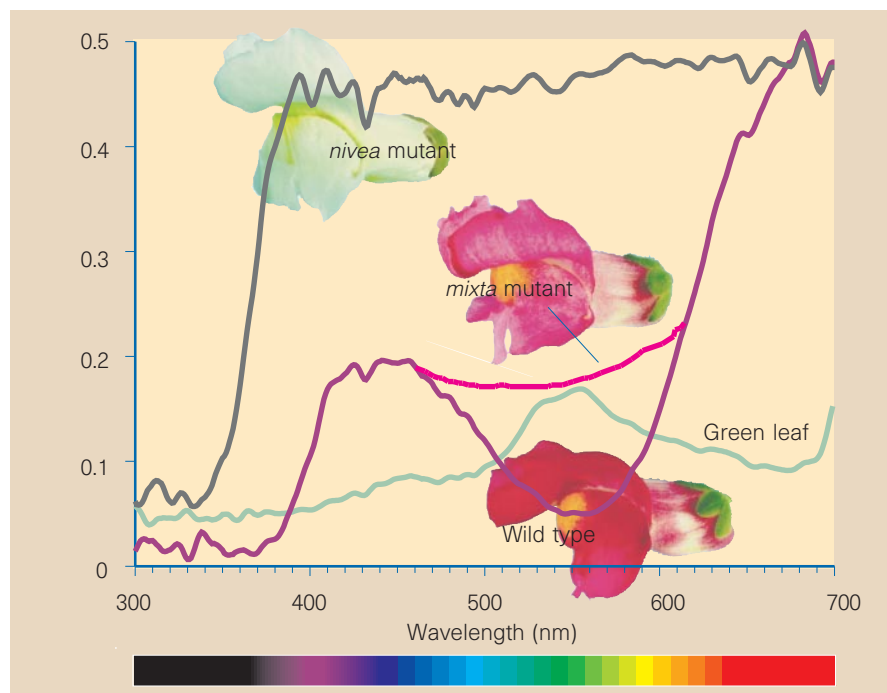


Figure 2 Reflectance spectra of *Antirrhinum majus*. A colour shift from purple to pink (as in *mixta* mutants) is generally mediated by an increase in green reflectance⁵ (pink curve). Ultraviolet-reflecting white flowers (such as the *nivea* mutants) usually reflect all light above 360 nm (grey line), whereas ultraviolet-absorbing white flowers cut off at 420 nm. Mutant spectra are inferred from flowers of over 1,000 other plant species⁵.

and shapes attract different pollinators and, indeed, that they often represent adaptations of plant species to specialize on different pollinators. This is the idea of 'pollination syndromes', which Glover and Martin refer to throughout, and which underlies other studies of floral evolution.

For example, the mapping by Bradshaw *et al.*⁹ of quantitative 'speciation genes' in monkey flowers, genus *Mimulus*, assumes that differences in flower colour and shape erect reproductive barriers between lineages, transforming them into separate biological species. This requires that different colours and shapes automatically confer specialization on different pollinators, but the evidence¹⁰ suggests otherwise — different insects and other pollinating animals possess colour vision with broad spectral sensitivity, and colour is mainly an advertisement rather than an innate attractant for experienced pollinators. In the case of monkey flowers, Sutherland and Vickery¹¹ had earlier shown that, from direct observations of pollinators, bumble-bees and hummingbirds do not specialize on different colours and shapes. So, we may have to look beyond pollinators to explain reproductive isolation between plant species¹².

Whether a bee or other pollinator prefers one flower over another is a result of many factors: innate preference (if any); past experience of one flower as more rewarding than the other; familiarity with one or the other flower; ease of handling of one or the other; and sensory limitations that may make one

flower more detectable. These components have been dissected for honey bees and other bees, but we are just beginning to merge these advances with pollination ecology, and to explore the implications for plant fitness. This merger is an exciting prospect — it should allow us to transform the traditional framework of pollination syndromes into a more satisfying picture of the interactions between flowers and pollinators, and their evolutionary consequences. □

Nick Waser is in the Department of Biology, University of California, Riverside, California 92521, USA.

e-mail: waser@citrus.ucr.edu

Lars Chittka is at Zoologie II, Biozentrum, Am Hubland, D-97074 Würzburg, Germany.

e-mail: chittka@biozentrum.uni-wuerzburg.de

- Glover, B. J. & Martin, C. *Heredity* **80**, 778–784 (1998).
- Levin, D. A. & Brack, E. T. *Evolution* **49**, 1017–1022 (1995).
- Fineblum, W. L. & Rausher, M. *Ecology* **78**, 1646–1654 (1997).
- Campbell, D. R., Waser, N. M. & Price, M. V. *Evolution* **48**, 55–68 (1994).
- Chittka, L., Shmida, A., Troje, N. & Menzel, R. *Vision Res.* **34**, 1489–1508 (1994).
- Dafni, A., Lehrer, M. & Kevan, P. G. *Biol. Rev.* **72**, 239–282 (1997).
- Menzel, R. in *Neurobiology of Comparative Cognition* (eds Kesner, R. P. & Olten, D. S.) 237–292 (Erlbaum Inc., Hillsdale, NJ, 1990).
- Giurfa, M., Núñez, J., Chittka, L. & Menzel, R. *J. Comp. Physiol. A* **177**, 247–259 (1995).
- Bradshaw, H. D. J., Wilbert, S. M., Otto, K. G. & Schemske, D. W. *Nature* **376**, 762–765 (1995).
- Chittka, L. & Waser, N. M. *Israel J. Plant Sci.* **45**, 169–183 (1997).
- Sutherland, S. D. & Vickery, R. K. *Great Basin Natur.* **53**, 107–117 (1993).
- Waser, N. M. *Oikos* **82**, 198–201 (1998).

Daedalus

Total protection

Immunization, that brilliant medical invention, is a way of arming the body's defences in advance of an attack. You inject a foreign protein characteristic of the disease organism; the immune system raises antibodies against it, and acquires expertise in detecting the protein and neutralizing its threat. Any pathogen bearing that protein is thereafter rapidly overwhelmed by primed and practised immunological defences.

Daedalus now wants to generalize this technique. The immune system seems to be able to produce antibodies to any protein, or any number of proteins. So DREADCO biologists are devising a universal immunization. At first they planned to exploit modern combinatorial chemistry to synthesize a mixture of all possible proteins, and inject it into test subjects. But the combinatorics defeated them. With 20 possible amino acids at each link in a protein chain, an immunization containing just one molecule of all possible 20-link chains would weigh well over a kilogram. And much longer chains are common.

However, life uses only a small subset of proteins, those with useful foldings; many of them are common to several species. So Daedalus's team is now scouring farms, zoos, botanical gardens, and culture archives to acquire samples from, or a specimen of, every known living thing. They plan to mix and homogenize the specimens, and extract their proteins. The resulting mixture will still be wildly complicated, but should be a feasible immunization.

DREADCO's 'Noah's Ark Vaccine' will be powerful indeed. An injection of it will bring a horrific feverish reaction as the subject's immune system is challenged simultaneously by millions of foreign proteins. A graded series of shots will be needed, starting with very tiny doses and only gradually building up to the full prophylactic dose. The subject will then be immune to absolutely everything.

Given to the whole population, Noah's Ark Vaccine will transform medicine. Colds, 'flu, infections, all the bacterial and viral diseases will vanish. Even if the baffled organisms mutate, they will simply hit another component of our fully primed immune systems. Even a pathogen protein not in the vaccine may still be close enough to one that is, to trigger its response. Medical costs will be cut right back. Only diseases of the immune system itself, such as AIDS, will remain to plague us.

David Jones

Modelled moons

How do you bring a flat picture to three-dimensional life? Early photographers met the same visual challenges that confronted Galileo. They used ingenious methods to build relief models of the lunar landscape.

Martin Kemp

The problem of how to throw telescopic observations literally into relief had beset the representation of celestial bodies since Galileo's time. Galileo's own accomplished watercolours of the irregular light and dark blotches on the Moon, which cast them decisively as the highlights and shadows of mountains and valleys, set the standards for articulate looking. The advent of astronomical photography re-posed such challenges to seeing and knowing.

The first surviving image of the Moon of real quality was exhibited by John Whipple and George Bond at the Great Exhibition in 1851, to be followed by rival 'Moonscapes' by other photographers, including Lewis Rutherfurd, whose detailed masterpiece of tonal modelling in 1865 did much to silence critics of the camera as an adjunct to astronomical observation.

The most sustained exploration of photography in the context of other visual techniques occurs in Nasmyth and Carpenter's photographically illustrated book of 1874, *The Moon: Considered as a Planet, a World, and a Satellite*. James Carpenter was an astronomer at the Royal Observatory in Greenwich; James Nasmyth was an engineer whose father, Alexander, was regarded as the 'father of Scottish landscape painting'.

Nasmyth and Carpenter used a notable range of diagrammatic and representational modelling. They used prints of Nasmyth's wonderfully crisp drawings, a photograph of the Moon by Warren De la Rue and Joseph Beck, a 'picture map' with some shading, a 'skeleton map' of lines alone and an imaginative impression of an 'ideal lunar landscape'. In addition they illustrated a representation of an eclipse of the Sun as viewed from the Moon, plaster models of the lunar surface photographed under immaculately controlled lighting, and a photograph of a laboratory simulation of a live volcano.

The authors claimed "upwards of thirty years of assiduous observation" during which "every favourable opportunity has been seized to educate the eye".

They resorted to exactly the same kind of visual analogies that had performed such a vital service for Galileo and the telescopists and Robert Hooke and the microscopists.

To this end, Nasmyth and Carpenter illustrate striking photographic 'heliotypes' of an aged hand and a shrivelled apple under lateral illumination.



"Photograph of the Moon" by Warren De la Rue and Joseph Beck and "Back of hand and wrinkled apple" (below). Both images are from *The Moon: Considered as a Planet, a World, and a Satellite* by James Nasmyth and James Carpenter, 1874.



Their visual analogies involve both appearance and process, in the case of the wrinkles between the mechanics of the folding that occurs when the inner core acquires a surface area smaller than its covering skin. Their methods of modelling brought a compelling sense of the Universe and its parts in 'perpetual mutation', evolving from primeval matter. This conviction is nicely

expressed on the first page of their text. Since "the same laws work in the great as well as the smallest processes of nature, we are compelled to believe in an antecedent state of existence of the matter that composes the host of the heavenly bodies". □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

The phylogeny of *The Canterbury Tales*

Geoffrey Chaucer's *The Canterbury Tales* survives in about 80 different manuscript versions¹. We have used the techniques of evolutionary biology to produce what is, in effect, a phylogenetic tree showing the relationships between 58 extant fifteenth-century manuscripts of "The Wife of Bath's Prologue" from *The Canterbury Tales*. We found that many of the manuscripts fall into separate groups sharing distinct ancestors.

Manuscripts such as these were created by copying, directly or indirectly, from the original material (written, in the case of *The Canterbury Tales*, in the late fourteenth century). In the process of copying, the scribes made (deliberately or otherwise) changes, which were themselves copied. Textual scholars have developed a system for reconstructing the relationships between textual traditions by analysing the distribution of these shared changes, and have constructed family trees (stemmata) on the basis of the results, with the ultimate aim of establishing precisely what the author actually wrote. This analysis is carried out manually and is feasible only for a few manuscripts of short texts. The sheer quantity of information in a tradition the size of *The Canterbury Tales* defeats any system of manual analysis.

However, the principle of historical reconstruction is similar to the computerized techniques used by evolutionary biologists to reconstruct phylogenetic trees of different organisms using sequence data. We therefore applied phylogenetic techniques to *The Canterbury Tales* using the 850 lines of 58 surviving fifteenth-century manuscripts of “The Wife of Bath’s Prologue”. We believe this to be the first full tradition of a major work to be analysed in this manner.

It may be inappropriate to impose a tree-like structure on such data sets, so we used the method of split decomposition implemented in the program SplitsTree², in addition to the cladistic analysis of PAUP³. Figure 1 shows a SplitsTree analysis of 44 of

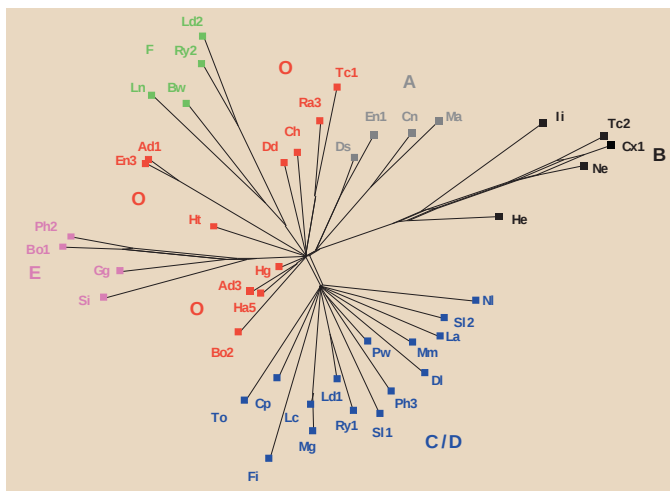


Figure 1 SplitsTree analysis of 44 manuscripts of "The Wife of Bath's Prologue" from Chaucer's *The Canterbury Tales*⁴. The two- or three-character codes indicate individual manuscripts, whereas the large capitals indicate groups of manuscripts, which are coloured the same.

the 58 manuscripts. Very similar results were given by PAUP (not shown). Several manuscripts form groups (A, B, C/D, E and F), each descended from a single and distinct common ancestor. The remaining 14 manuscripts were removed from the analysis shown in Fig. 1, as they were likely to have been copied from more than one exemplar, either by deliberate conflation of readings or by changing the exemplar during the course of copying. These manuscripts were identified by comparison of the trees generated with different regions of the text, which showed that their position in the analysis varied dramatically depending on which region was used. The central point is likely to represent the ancestor of the whole tradition. The manuscripts grouped as O are particularly crucial; their position near to the centre suggests that they all descend from Chaucer's original, and may therefore contain crucial evidence about this original. However, most of them have been ignored by scholars.

From this analysis and other evidence, we deduce that the ancestor of the whole tradition, Chaucer's own copy, was not a finished or fair copy, but a working draft containing (for example) Chaucer's own

notes of passages to be deleted or added, and alternative drafts of sections. In time, this may lead editors to produce a radically different text of *The Canterbury Tales*. These results also demonstrate the power of applying phylogenetic techniques, and particularly split decomposition, to the study of large numbers of different versions of sizeable texts.

Adrian C. Barbrook, Christopher J. Howe
Department of Biochemistry,
University of Cambridge,
Tennis Court Road, Cambridge CB2 1QW, UK
e-mail: c.j.howe@bioc.cam.ac.uk

Norman Blake
Humanities Research Institute,
Arts Tower, University of Sheffield,
Sheffield S10 2TN, UK

Peter Robinson
The Division of Learning Development,
Gateway House, De Montfort University,
Leicester LE1 5XY, UK

1. Blake, N. F. *The Textual Tradition of The Canterbury Tales* (Edward Arnold, London, 1985).
2. Huson, D. H. *Bioinformatics* 14, 68–73 (1998).
3. Swofford, D. L. *PAUP Version 3.1.1*. (Smithsonian Institute, Washington DC, 1993).
4. Robinson, P. M. W. in *The Canterbury Tales Project: Occasional Papers Vol. II* (eds Blake, N. F. & Robinson, P. M. W.), 69–132 (Office for Humanities Communication, London, 1997).

Green-wave phenology

Phenology, the traditional study of seasonal plant and animal activity driven by environmental factors, has found new relevance in research into global climate change^{1,4}. Global phenology research so far has concentrated on measurements obtained by satellites, downplaying connections of these measures to information obtained on the surface^{1,4}, perhaps because of a lack of conventional, surface-based phenological data. However, an integration of conventional

and satellite-derived measures is needed to understand better the mid-latitude spring onset of photosynthesis, known as the ‘green wave’ or ‘green-up’^{5–7}. Here I show how a surface-based green-wave model can extend the monitoring of climatic variability back to 1900, providing a longer-term context for the more limited, recent data obtained from satellites.

Simple events, such as flower blooms or insect hatchings, offer clues to the workings of complex processes⁸. The spring green wave signals the start of the growing season and profoundly influences subsequent

productivity⁷; plant activity is more responsive to spring weather than to other seasons. Thus, indices that measure the onset of the green wave are ideal biological measures of climatic variability. Despite these attributes, traditional phenological studies are typically local-scale agricultural projects, rarely involving ecosystem- or global-scale processes. In contrast, satellite remote sensing has been widely used for biospheric monitoring, including for phenology^{1,2,4,9}.

The daily global coverage by satellites makes them essential to investigations of green-up. However, remote-sensing data

extend back only to about 1982, are limited by cloud cover, and may not distinguish between overstorey (the dominant trees in the forest) and understorey (shrubs and herbaceous plants on or near the forest floor) phenology in forests^{2,5}.

Other biospheric phenomena have been assessed globally through strategic combinations of satellite and surface measurements³. The above limitations of remote-sensing 'phenology' highlight the complexity of the green wave, and suggest the need for a similar strategy here. What the satellite senses and what is observed on the ground are both integral parts of the green wave. Conventional phenological data — carefully selected according to species, and globally distributed — should play a crucial role in global green-wave research. As global-scale phenological networks do not yet exist, empirical green-wave models can partly fill the void.

Models that simulate phenology using meteorological data offer several advantages, if based on appropriate plants^{10,11}. They allow reconstruction of green waves back through the period for which instrumental records of weather are available, providing a context for short-interval satellite data. Also, models serve as 'anchor points', binding together commonalities among records for native species in adjacent biomes and remote-sensing observations. Empirical models (spring indices) can closely mimic actual plant first-leaf and first-bloom events, correlate with native-species data, and reveal changes in lower-atmospheric processes at the ecosystem scale^{5,6,10,12,13}, illustrating their potential use in longer-term studies (Fig. 1).

As an example, I studied green-wave changes in eastern North America⁵ from 1900 to 1995 (Fig. 2). Meteorological data were from the Historical Climatology

Network Daily Temperature and Precipitation Data (CDIAC, Oak Ridge National Laboratory). I calculated several indices for each year for all stations over their respective periods of record. All indices showed considerable year-to-year variation, without striking long-term trends, but with significant shorter-period changes. For example, from 1978 to 1990, first-leaf spring index dates became earlier at a rate of about 1 day per year (trend adjusted $r^2 = 0.43$, α -level = 0.009; Fig. 2).

The tendency of global phenology research to concentrate on satellite-based measurements makes this approach fundamentally incomplete. Satellite measurements provide broad areal coverage but reveal only one aspect of green-up. All measures — empirical models, native-species phenology, and appropriately calibrated satellite indices — need to be understood and interconnected for maximum effectiveness^{5,6}.

Global-scale plant phenology networks should be established to strengthen this research strategy. However, empirical models offer a way to reconstruct past green waves, and allow regional comparisons. Although imperfect and of limited geographical application, these models provide one of the few independent comparisons available for satellite measurements of plant activity⁹. Connections between conventional phenology and remotely sensed greenness measures at the ecosystem level are being confirmed, and further studies are underway⁴. It is encouraging that a satellite-derived change in green-up⁹ (earlier by 8 ± 3 days over the 1981–91 period) is consistent with the trend, over roughly the same period, that I describe here (ten days earlier during the 1980–90 period, inferred from the 1978–90 regression line). But this short period appears unremarkable when compared with the overall first-leaf spring

index variability in eastern North America (Fig. 2).

My results endorse empirical models as useful partners for satellite-derived green-wave measures. Studies of other regions, especially those with long-term warming trends, can provide a context for evaluating satellite measurements that indicate earlier greenness onset over large areas. Energy budget analyses coupled with surface and satellite phenology (for example, examination of phenological effects on the exchange of latent and sensible heat between the surface and the atmosphere) should also prove rewarding in further understanding and monitoring spring plant–climate interactions.

Mark D. Schwartz

Department of Geography,
University of Wisconsin–Milwaukee,
Milwaukee, Wisconsin 53201-0413, USA
e-mail: mds@uwm.edu

1. Moulin, S., Kergoat, L., Viovy, N. & Dedieu, G. *J. Clim.* **10**, 1154–1170 (1997).
2. Reed, B. C. *et al.* *J. Veg. Sci.* **5**, 703–714 (1994).
3. Sellers, P. J. *et al.* *J. Clim.* **9**, 706–737 (1996).
4. White, M. A., Thornton, P. E. & Running, S. W. *Glob. Biogeochem. Cycles* **11**, 217–234 (1997).
5. Schwartz, M. D. in *Phenology of Seasonal Climates* (eds Lieth, H. & Schwartz, M. D.) 23–38 (Backhuys, Netherlands, 1997).
6. Schwartz, M. D. *Int. J. Biometeorol.* **38**, 18–22 (1994).
7. Running, S. W. & Hunt, E. R. in *Scaling Physiological Processes, Leaf to Globe* (eds Field, C. & Ehleringer, J.) 144–157 (Academic, New York, 1993).
8. Leopold, A. & Jones, E. *Ecol. Monogr.* **17**, 81–122 (1947).
9. Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. *Nature* **386**, 698–702 (1997).
10. Schwartz, M. D. *Phys. Geogr.* **14**, 536–550 (1993).
11. Hopp, R. J. & Vittum, M. T. *Organic Gard. Farm.* **24**, 127–129 (1977).
12. Schwartz, M. D. *J. Clim.* **9**, 803–808 (1996).
13. Schwartz, M. D. *Month. Weather Rev.* **120**, 2570–2578 (1992).

A posteriori teleportation

The article by Bouwmeester *et al.*¹ on experimental quantum teleportation constitutes an important advance in the burgeoning field of quantum information. The experiment was motivated by the proposal of Bennett *et al.*² in which an unknown quantum state is 'teleported' by Alice to Bob. As illustrated in Fig. 1, in the implementation of this procedure by Bouwmeester *et al.*¹, an input quantum state is 'disembodied' into quantum and classical components, as in the original protocol². However, in contrast to the original scheme, Bouwmeester *et al.*'s procedure necessarily destroys the state at Bob's receiving terminal, so a 'teleported' state can never emerge as a freely propagating state for subsequent examination or exploitation. In fact, teleportation is achieved only as a postdiction.

Bouwmeester *et al.* used parametric down-conversion from two sources (SI, SII in Fig. 1) in an attempt to teleport the

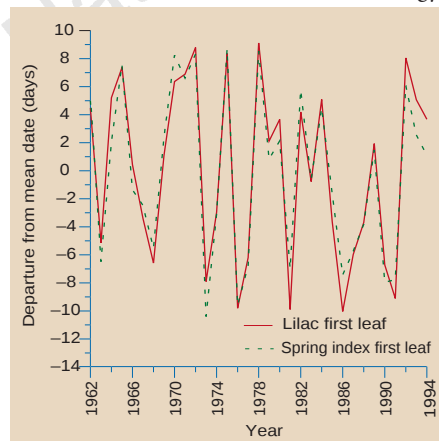


Figure 1 Comparison of annual departures from the mean of spring index first-leaf dates (derived from an empirical model) and of *Syringa chinensis* 'red rothomagensis' (lilac) first-leaf dates (derived from actual data), 1962–94 (details of the network of the 183 stations with relevant data, and methods, are available from the author).

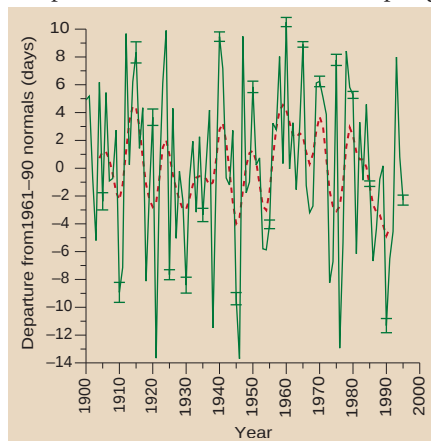
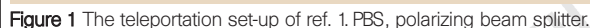


Figure 2 Eastern North American⁵ departures from the mean of spring index first-leaf dates, 1900–95, with ± 1 standard error bars, and smoothed trend produced by a nine-year, moving average, normal curve filter (dotted line) (details of the network of the 465 relevant stations, and methods, are available from the author).



Under relaxed conditions, requiring only threefold coincidence of detectors p and (f1, f2), teleportation is achieved when the fields of beams 1 and 3 match with sufficiently high fidelity. In the simplest approximation, type II parametric down-conversion of modes (i, j) generates wavepacket states as follows:

where A_0 , A_1 and A_2 are the coefficients for obtaining no (vacuum), one and two down-converted pairs, respectively, and $(i, j) = (1, 4) (2, 3)$. Of these terms, only states corresponding to the second term are selected by fourfold coincidence, as specified by equations (2) and (3) of ref. 1. However, anything less than complete destruction of the output 3 necessarily leaves undesirable terms that reduce F .

field 4 at detector p, which projects the field in beam 1 accordingly. Joint detection at (f1, f2) then provides threefold coincidence with p, yielding a statistical mixture for the field 3 arriving at Bob's station. The fraction of the state $|\Phi\rangle$ in this mixture gives F . To the lowest order in the down-converter coupling strength, the Bouwmeester *et al.* scheme yields a 50:50 mixture of the vacuum state $|0\rangle$ and the desired state $|\Phi\rangle$, with $F=1/2$, so that there is never a physical state with high teleportation fidelity. Indeed, Bob could achieve this same fidelity, $F=1/2$, by abandoning teleportation altogether and transmitting randomly selected polarization states. Faced with this state of affairs, the experiment of ref. 1 obtains a surrogate for high fidelity by destructively recording the field 3 at (d1, d2).

We emphasize that the nature of the mixture containing the vacuum state has definite physical implications, which can be verified by more general measurements than photon counting (for example, by quantum-state tomography). Moreover, the freedom of a potential consumer of the output from Bob's receiving station to select alternative detection strategies means that classical analogies fail.

(for example, by cascading conventional detectors) could provide an effective remedy. Appropriate selection could be implemented with a polarization-independent quantum non-demolition measurement of the total photon number at Bob's end. Alternatively, pre-selection could be implemented by enhancing the coupling between modes (2, 3) relative to modes (1, 4).

Samuel L. Braunstein

H. J. Kimble

1. Bouwmeester, D. *et al.* *Nature* **390**, 575–579 (1997).

2. Bennett, C. H. *et al. Phys. Rev. Lett.* **70**, 1895–1899 (1993).

Bouwmeester et al. reply — Braunstein and Kimble observe correctly that, in the Innsbruck experiment, one does not always observe a teleported photon conditioned on a coincidence recording at the Bell-state analyser. In their opinion, this affects the fidelity of the experiment, but we believe, in contrast, that it has no significance, and that when a teleported photon appears, it has all the properties required by the teleportation protocol. These properties can never be achieved by “abandoning teleportation altogether and transmitting randomly selected polarization states” as Braunstein and Kimble suggest. The fact that there will be events where no teleported photons are created merely affects the efficiency of the experiment. This suggests that the measure of fidelity used by Braunstein and Kimble is unsuitable for our experiment.

During the detection of the teleported photons, no selection was performed based on the properties of these photons. Therefore, no *a posteriori* measurement in the usual sense as a selective measurement was performed. The detection of the teleported photon could have been avoided altogether if we had used a more expensive detector, p, that could distinguish between one- and two-photon absorption. The inability of our teleportation experiment to perform such refined detections does not, however, imply that “a teleported state can never emerge as a freely propagating state...”. Braunstein and Kimble do not, therefore, reveal a principal flaw in our teleportation procedure, but merely address a non-trivial practical question.

NATURE | VOL 394 | 27 AUGUST 1998

DNA methylation models histone acetylation

One of the main determinants of chromatin structure is histone acetylation¹. Local chromosomal acetylation can be regulated dynamically, both through the involvement of transactivating factors with intrinsic histone acetylase activity, and through the recruitment of deacetylase complexes that repress gene expression². Histone acetylation status is transiently modified from one state to another in response to physiological changes operating in the cell. It is not yet known, however, how the basic histone acetylation profiles on tissue-specific and housekeeping gene sequences are established during normal development. Here we show that DNA methylation takes part in this process by inducing decreased levels of chromatin acetylation.

Immunoprecipitation was used to isolate specifically the highly acetylated mononucleosome fraction¹ from stably transfected cells carrying either methylated or unmethylated copies of the thymidine kinase gene (*tk*) from herpes simplex virus (HSV). This technique clearly separates the transcriptionally active acetylated nucleosome fraction from bulk chromatin. The constitutively expressed β -actin gene, for example, is enriched in the acetylated fraction (4.5-fold), as are exogenous copies of

the unmethylated *tk* gene (also 4.5-fold). In contrast, methylated *tk* sequences behave like endogenous inactive genes (for example, the *Ig κ* gene), and are not preferentially associated with acetylated histones (Fig. 1a). Because identical DNA sequences were used in both transfection experiments, this decreased level of chromatin acetylation must be due to the presence of 5-methyl cytosine. These experiments therefore demonstrate that DNA methylation influences local histone acetylation, and it has now been shown that the methyl-binding protein MeCP2 is probably involved in this process by recruiting histone deacetylases^{3,4}.

We investigated whether this under-acetylation contributes to transcriptional inhibition. We used the drug trichostatin A (TSA) to increase general histone acetylation in the cell⁵ in an attempt to bypass the inhibitory effects of DNA methylation (Fig. 1b). Treatment with TSA resulted in a 2.5-fold increase in transcription from a fully methylated chicken *tk* gene. This stimulation was the result of a specific release from methyl-induced repression, rather than a boost in gene expression, as transcription from both the endogenous β -actin and the exogenous unmethylated chicken *tk* genes was unaffected by TSA treatment. Maximal induction by TSA (7-fold) was obtained on an HSV *tk* gene methylated exclusively in non-promoter regions⁶. Most DNA methyl sites in the genome are not located within regulatory elements. Thus, histone deacetylation may be an important

mediator of the global repression associated with DNA methylation.

It has been shown that the presence of methyl groups on exogenously introduced sequences makes local chromatin structure less accessible to the transcription machinery⁷ and molecular probes, such as DNaseI⁸. To determine whether this inactive chromatin structure is brought about by histone deacetylation, cells carrying fully methylated copies of the HSV *tk* gene were treated with TSA. Although remaining fully methylated (data not shown), these sequences are converted to a DNaseI-sensitive conformation (Fig. 1c), indicating that histone deacetylation itself is a critical factor in the generation of DNaseI-insensitive structures associated with methylated DNA.

The mammalian genome has a bimodal methylation pattern in which CpG islands in housekeeping genes are constitutively unmodified, whereas most tissue-specific genes are generally methylated in every cell type except those that actually express the gene⁹. This profile is established through a series of *de novo* and demethylation events, and is perpetuated by maintenance methylation through subsequent cell divisions¹⁰. We propose that this pattern helps to set up local histone acetylation states. After replication, newly deposited histones in unmethylated regions may undergo acetylation, but nucleosomes containing methylated DNA will be subject to deacetylation throughout the cell cycle.

Histone acetylation is a fundamental regulatory mechanism for controlling gene accessibility, and it is therefore not surprising that the cell uses a variety of different molecular pathways for its modulation. Our results indicate that in higher organisms DNA methylation may serve as a unique mechanism for setting up local histone deacetylation, and so generate maintainable epigenetic chromosomal states.

S. Eden, T. Hashimshony, I. Keshet, H. Cedar

Department of Cellular Biochemistry, Hebrew University, Ein Kerem, Jerusalem 91120, Israel
e-mail: cedar@md2.huji.ac.il

A. W. Thorne

Biophysics Laboratories, University of Portsmouth, Portsmouth PO1 2DT, UK

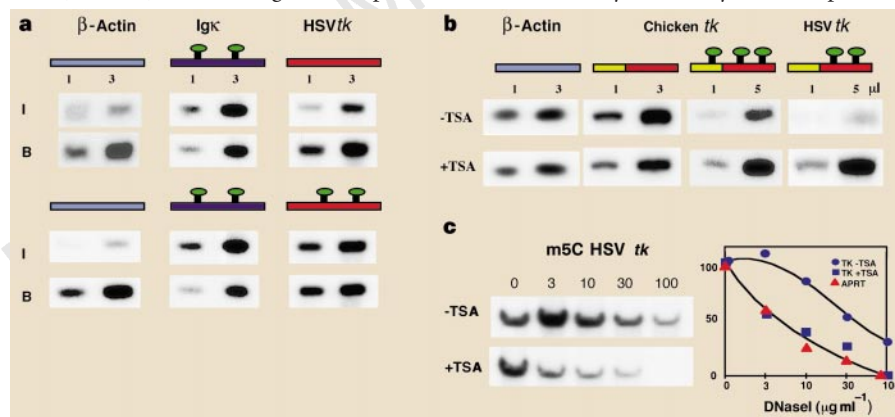


Figure 1 DNA methylation, histone acetylation and chromatin structure. **a**, Micrococcal nuclease-prepared mononucleosomes from pools of L-cell colonies stably transfected with unmethylated (upper portion) or *in vitro* S1-methylated (lower portion) HSV *tk* gene were precipitated with antibodies mainly specific to acetylated histone H4 (ref. 1). DNA was extracted from input (I) and bound (B) fractions and subject to quantitative polymerase chain reaction (PCR) using primers for HSV *tk* and mouse β -actin or *Ig κ* . Two concentrations (1- and 3-fold) are shown. The degree of enrichment was determined by phosphorimager or gel scan analysis. Methylated HSV *tk* is present in higher average copy number than the unmethylated vector. Green blobs indicate DNA methylation. **b**, Cells were stably transfected with unmethylated or *in vitro* S1-methylated chicken *tk* gene or the hybrid HSV *tk* gene⁶ methylated in the coding region (red) but not the promoter (yellow). Quantitative PCR conditions were established empirically for each gene and cell type, and carried out on two concentrations (1- and 3-fold, or 1- and 5-fold) of equivalent amounts of cDNA from untreated or TSA-treated (50 ng ml⁻¹, 24 h) cells. The β -actin results are from one of the cell lines. **c**, Nuclei from a mixture of cells containing either the fully methylated HSV *tk* or the unmethylated hamster *aprt* gene ($\pm$ TSA) were digested with increasing concentrations of DNaseI and the DNA was extracted and subject to blot hybridization analysis⁸. The graph (based on phosphorimaging of the blots) shows *tk* sensitivity to DNaseI (percentage of total DNA) as compared with *aprt* (average of $\pm$ TSA samples, which were very similar) on the same blot.

- Hebbes, T. R. et al. *EMBO J.* **13**, 1823–1830 (1994).
- Grunstein, M. *Nature* **389**, 348–352 (1997).
- Nan, X. et al. *Nature* **393**, 386–389 (1998).
- Jones, P. L. et al. *Nature Genet.* **19**, 187–191 (1998).
- Yoshida, M., Horinouchi, S. & Beppu, T. *Bioassays* **17**, 423–430 (1995).
- Keshet, I., Yisraeli, J. & Cedar, H. *Proc. Natl Acad. Sci. USA* **82**, 2560–2564 (1985).
- Buschhausen, B., Wittig, M., Graessmann, M. & Graessmann, A. *Proc. Natl Acad. Sci. USA* **84**, 1177–1181 (1987).
- Keshet, I., Lieman-Hurwitz, J. & Cedar, H. *Cell* **44**, 535–543 (1986).
- Yeivin, A. & Razin, A. in *DNA Methylation: Molecular Biology and Biological Significance* (eds Jost, J. P. & Saluz, H. P.) 523–568 (Birkhauser, Basel, 1993).
- Kafri, T. et al. *Genes Dev.* **6**, 705–714 (1992).

Anyone for tenets?

Great Feuds in Science

by Hal Hellman

John Wiley & Sons: 1998. 256pp. £16.50, \$24.95

Walter Gratzer

There are no sects in geometry, according to Voltaire, but he reckoned without human folly and the vanity of scholars. Pythagoras, who cleaved passionately to a belief in the universality of rational numbers, was said to have been so incensed when his disciple proved $\sqrt{2}$ at least to be an exception that he had the young man killed. This might strike us today as putting too high a price on conjecture, yet well within living memory public endorsement of Mendelian genetics was enough to earn Russian biologists a one-way trip to the Gulag; and Heisenberg's affirmation of quantum mechanics came close to doing for him in 1937.

Hal Hellman has assembled a series of entertaining tales about clashes of ideas or more often of personalities. He kicks off with Galileo, wrestling in vain with Catholic orthodoxy and a Pope already beleaguered by politics. Galileo's *Dialogue concerning Two World Systems* remained on the *Index* of forbidden books until 1822, and the sentence on its author was not annulled until 1992, when

the Holy Mother Church finally threw in the towel and conceded that the Earth goes round the Sun after all.

Hellman gives only T. H. Huxley's version of his famous encounter with Bishop Wilberforce ("The Lord has delivered him into my hands!"). J. D. Hooker wrote the next day to Darwin that Huxley had been ineffectual and it was he himself who had delivered the decisive blow. In any case, by this time it was not that daring to take on a bishop, especially an ageing Tory like the sententious Soapy Sam. Huxley, of course, loved polemic; it was for him, he confessed, "like gin to a reformed drunkard", but witnesses thought that his mind was so clouded by rage at Wilberforce's casuistry as to rob him of coherent speech.

You do not, as Hellman's narratives repeatedly show, have to be right to win an argument. He cites many fine examples of heady invective without parallel in our time. Thomas Hobbes, like most intellectuals of his day, was an amateur mathematician. Everyone in refined society, it seems, was expected to take a position on such issues as squaring the circle, and Hellman has even discovered the sad case of a suitor rejected by his innamorata because of the poverty of his views on the subject. Hobbes, however, came

up against the real thing in the form of the formidable John Wallis, mathematician and follower of Newton. Here is how Hobbes seeks to settle the argument on the same problem. Having referred to Wallis's "scurvy book", packed with "mere ignorance and gibberish", he delivers this stirring peroration: "So go your ways, you Uncivil Ecclesiastics, Dedoctors of morality, Unasinous Colleagues, Egregious pair of Issachars, most wretched Vindices and Indices Academia-trum" (Hellman translates). But injurious men, as has been said, brook no injuries, and so the argument swayed back and forth with unrelenting acrimony.

The disagreeable Isaac Newton also had a fine line in vituperation and subterfuge, which included an enthusiastic commendation of his own work, written anonymously by himself under the imprint of the Royal Society. His most persistent animus was for Leibniz. Neither had been in any great haste to publish their invention of the calculus: Leibniz waited nine years and Newton forty. But news leaked out and accusations of plagiarism flew. Chauvinism played its part and so did theological differences, for both men were fervently religious.

Newton was troubled by the implication in his laws of motion that the universal clock-



Battling for the centre: an anonymous depiction of Galileo's trial in 1632.

ANONYMOUS/THE BRIDGEMAN ART GALLERY

work, once wound up by God, would proceed for ever in a deterministic way, resistant therefore to intervention by man through prayer. He hoped that perhaps the residual irregularities in planetary motion would in time accumulate and disrupt the clockwork, when God would surely intervene to set it right. Leibniz regarded such concerns as absurd and a slight on the perfection of God's works. (It was Leibniz's compatriot, the physicist and aphorist Liechtenstein, who spoke of "God, who winds our sundials".)

Newton, in the eyes of the world, came out the winner and exulted that he had broken Leibniz's heart. So Newton finished up in Westminster Abbey, while Leibniz, rejected in favour of Newton even by his sometime patron, George I, died unregarded and in poverty, his secretary the sole mourner at the funeral.

Voltaire, another ferocious polemicist, took a stand in the long-running embryogenesis controversy: were all embryos of a human lineage created by God, like an infinite succession of nested Russian dolls, requiring merely the impulse of conception to bring forth the next, or were they generated by elements from father and mother, as the resemblance of a child to both its parents implied? This question was linked to the broader debate on whether creation was continuous. Voltaire's opponent was the respected biologist and cleric, John Needham, who devised an experiment to test for spontaneous generation of life: when a mutton broth, sterilized by heating in a sealed vessel, was examined some days later, it was found to be seething with microscopic life.

Voltaire had already seen off his compatriot, Maupertuis (the man who, among other brainwaves, conceived of bringing up a child in total isolation to establish whether Latin or Greek was the primal language), in a series of venomous lampoons on the same subject. He now launched himself at Needham, whom he denounced as an Irish Jesuit (which he was not) and a homosexual to boot. Needham replied in kind. Buffon, who admired Needham, also became a target of Voltaire's scorn, but responded with spirit: Voltaire's jealousy of others' success, he wrote, had "intensified his bile, which has been cooked with age, so that he seems to have formed a project intending to bury all of his contemporaries while they are living". The debate was settled by the great biologist, the Abbé Spallanzani, who anticipated Pasteur by repeating Needham's experiment with properly sealed and sterilized flasks.

In a curious coda, Hellman makes a number of disconcerting assertions. Even, he opines, if the "preformationists" have been vanquished, the vitalist position that living matter differs in some fundamental way from inert remains tenable. He also wonders what Voltaire would say if "for example, cancer is found to be caused by viruses or even

some lower life form, created *de novo*". Such *obiter dicta*, scattered around the text, slightly undermine one's confidence in Hellman as a guide to matters of science.

Moving towards the twentieth century, Hellman lays bare for us some egos like malignant tumours. The two palaeontologists and dinosaur hunters Edward Cope, from the University of Pennsylvania, and Othniel Charles Marsh of Yale were divided less by differences of scholarly opinion than by corrosive personal jealousies and tussles over priority.

Hellman cites an analysis by Robert Merton, which concludes that whereas, in the eighteenth century, 92 per cent of simultaneous discoveries ended in disputes, the figure fell to 59 per cent in the latter half of the nineteenth and 33 per cent in the first half of the twentieth century; it still seems an outrageous proportion. In the case of Cope and Marsh, and to an extent that of Richard Leakey and Donald Johanson, digging in Africa for evidence of human origins, scientific questions were obscured by rivalry, rooted in barely concealed personal loathing.

Politics, too, can intrude, especially in the soft sciences, for, as Hellman shows, whether one embraces the conclusions of Margaret Mead on the nature of human relationships, or those of her nemesis Derek Freeman, tends to depend more on ideological inclination than on weighing a heap of elusive and disputed evidence.

We live in a time when it has become dangerous to make enemies of those who may tomorrow be reviewing one's paper for *Nature* or one's grant application. So scientific debate has become muted and decorous. The gaiety of the community is seldom enlivened any more by a vulgar brawl, but rivalries nurtured in private can be just as insistent.

If, as Hume put it, pride and hatred invigorate the soul (while love and humility enfeeble it), they may equally act as the engines of scientific progress. Gore Vidal put it in a nutshell: "Every time a friend succeeds a little something in me dies." □

Walter Gratzer is at the Randall Institute, King's College London, London WC2B 5RL, UK.

How to avoid making a fortune in medicine

The Story of Interferon: The Ups and Downs in the Life of a Scientist

by Kari Cantell

World Scientific: 1998. 239pp. \$22, £15

Jean Lindenmann

The style of biomedical research has changed drastically over the past 40 years. This is well illustrated by the fate of interferon. Beginning as a rather esoteric laboratory phenom-



Coiled to strike: interferon alpha.

enon (some would say laboratory artefact), it passed through a sort of cottage industry to end up part of the metropolis of modern biotechnology.

Kari Cantell has devoted his entire scientific life to making it available for evaluation as a drug. Thanks to his efforts, Finland was for many years the only place where clinically useful quantities of reasonably pure interferon could be obtained.

By education an MD, Cantell was soon attracted to medical research, and his first assignment already had some of the features that were to characterize his science. In Finland, a sparsely populated country where many children escaped infection with mumps, military training of fresh recruits resulted in alarmingly high rates of mumps in the barracks, often complicated by orchitis. So it was decided to prepare a vaccine from virus grown in chicken's eggs, to be administered to the entire Finnish army. In trying his best to improve the yields of virus, young Cantell thought that the more virus he injected into eggs, the more would come out — but this was not so. Cantell had run into a form of 'auto-interference'. From that moment he was hooked.

After postdoctoral training in Philadelphia with Kurt Paucker (a German Jew who had survived the Holocaust by escaping to Switzerland), at the laboratory where the Henles had done pioneering work on interference between viruses, Cantell returned to Finland as head of the virus laboratory at the State Serum Institute. There he stubbornly pursued the project of manufacturing human interferon from white blood cells in sufficiently large quantities for clinical trials. This work, in collaboration with Hans Strander, culminated in the first use of interferon as a treatment for cancer. The promising results with osteosarcoma (in compari-

son to historical controls) led to enormous hype in the media, which did a lot of harm in creating false hopes, but also mobilized money.

The US National Cancer Institute allotted funds for the purchase of Finnish interferon (now being produced, using Cantell's method, by the Finnish Red Cross). In 1980, the American Cancer Society ordered \$1 million worth of Finnish interferon, but its president decided to spend only one third of this sum on Finnish interferon and the rest on interferon made by an American company at double the price (also using methods learnt from Cantell). The significance of this is: that company had been founded by the president of the American Cancer Society. In spite of this rather blatant conflict of interests, Cantell, always very gentle in his judgement of others, finds excuses for the behaviour of this man.

During all these years, molecular biologists had been busy sharpening their tools. Derek Burke, in an article in *Scientific American*, had hinted at the possibility of producing large quantities of interferon in bacteria. Cantell was approached by Charles Weissmann to help in such a project. Messenger RNA from Cantell's interferon-secreting white blood cells served as the starting material for the production of interferon complementary DNA. This was successfully introduced into bacteria, which would use its machinery to start producing interferon in quantity.

Cantell never took patents on his procedure, and the interferon he manufactured himself at the State Serum Institute he gave away for clinical studies free of charge. Since his work was dependent on blood given voluntarily by Finnish donors, he felt that he should not make money out of it. Not all have been so open-handed, but Cantell writes, "it does not worry me that others have made fortunes as the result of our efforts".

Cantell's book is not a scholarly piece of historical writing. It has no bibliography, and no name index, which is a pity, because it gives information about many people which is otherwise difficult to find. It is based on his personal archive and his recollections, and so is likely to be useful to future historians of science as a primary source. Its factual content is, as far as I can tell, entirely accurate.

His concept of how science proceeds is positivistic, as indicated by the following statement: "Scientific phenomena are there already in nature, and just await discovery." This will not recommend him to modern sociologists of science, but even they will find some grist to their mill.

Commenting on a time when there appeared to be no useful applications for interferon, Cantell writes: "When the bubble seemed to burst, some of these companies no doubt came to regret their interest. But by then they had invested so much resource in

interferon that it became difficult to abandon further development." □

Jean Lindenmann is at Obere Geerenstrasse 34, CH-8044 Gockhausen, Switzerland.

The king of rocking roles

Annals of the Former World

by John McPhee

Farrar, Straus and Giroux: 1998. 695pp. \$35

Robert Muir-Wood

In 1978, John McPhee, a prolific and stylish non-fiction writer living in Princeton, New Jersey, decided to sample geology. The original intention was a short "Talk of the Town" piece for *The New Yorker*, based on an early morning encounter with a postgraduate geologist reading the history in a road cutting at the end of the George Washington Bridge. As the horns of passing trucks blared, they discussed spreading ridges and igneous petrology.

Inspired by this brief encounter, McPhee set out on a year-long journey across America. As he travelled, he was accompanied by a relay of four geologists, whose field areas and specialisms were explored in four short books, originally serialized in *The New Yorker*. The first, *Basin and Range*, which appeared in 1981, focused on Nevada; the second, *In Suspect Terrain* (1983), recorded journeys in the Appalachians; *Rising from the Plains* (1986) was set in Wyoming; to be followed in 1990 by *Assembling California*. A fifth and final chapter, "Crossing the Craton", set in Nebraska and Colorado, is now collected with all the earlier books in *Annals of the Former World*.

Each book was structured around journeys in the company of a single geologist, a

guide critical to the project and to the voice of each book. McPhee's original mentor and partner for *Basin and Range*, Kenneth Deffeyes, was an inspired choice. Deffeyes taught the introductory "Rocks for Jocks" course at Princeton University, which aims to seduce students into majoring in geology. Deffeyes's rich imagination is illustrated by one of many stories McPhee relates: immersed in his Archimedean bath-tub one day, thinking about the viscosity of the Earth's mantle, "suddenly he stood up and reached for a towel: 'Piano wire,' he said to himself, 'same viscosity as the mantle — 10^{22} poises — what keeps the piano tuner in business.'"

McPhee had struck a rich seam, and for several reasons. *Basin and Range* was in its time revolutionary (and a best seller), with a wit and vision that had never before been exercised in writing about the Earth.

McPhee had the exhilaration of an anthropologist discovering the tribe of geologists and their discipline for the first time. He viewed their taxonomy and terminology like a native language, and their fascination with excursions into the wilderness like a religion. He saw and described their ingenuity in reconstructing prehistory, and used the rocks as a foundation on which to recount the history of North American exploration, braided with the personality and roots of his companion.

McPhee also had an excellent eye for the contradiction between the lyrical — "Geologists in their closed conversations inhabit scenes that no one ever saw, scenes of global sweep, gone and gone again..." — and the more prosaic ironies of roadside geology — "Geologists on the whole are inconsistent drivers. When a road cut presents itself they tend to lurch and weave." Beyond the pleasure of the text — the "only country in Europe that is upside down is San Marino";



On the road: Nevada's Valley of Fire.

TONY STONE IMAGES

"Golf courses copy recently deglaciated topography in Scotland"; "Natural gas is to oil as politicians are to statesmen" — McPhee got to the heart of what four geologists thought in the early days of plate tectonics. The subplot to all the earliest books, and in particular *Basin and Range*, was the arrival of plate theory. The second book, *In Suspect Terrain*, is inadvertently the most revealing about this. The book's guide is Anita Harris, a field geologist with the US Geological Survey, who argues: "I get all heated up when some sweet young thing with three geology courses tells me about global tectonics, never having gone on a field trip to look at a rock"; "with the plate-tectonic model anybody can write a history of an area without having been there".

The books succeed or fail on the back of their main personality, which is why *In Suspect Terrain* is the weakest of the original four (as one tires of the bias) and *Assembling California* is perhaps the best, in the company of the erudite and charming Eldridge Moores, founder of the journal *Geology*, who was born in a tiny mining town in Arizona.

As the series continued, McPhee began to lose something of the innocence that inspired *Basin and Range*, and with it some of his ear for the unfamiliar. No longer will he be reading the road-cutting graffiti as his companion studies the rocks. Increasingly, he comes to sound like a geologist, quoting arcane terminology, lists of world localities where ophiolites of a particular class can be viewed, and even sets of papers and authors.

The shift is also evident in the production of this collection. The additional chapter on the Craton is intended to "complete" the geology of North America. The project now has ambitions to be a more comprehensive textbook, and is accompanied by an introduction that provides signposts on where in the text to find the key elements of geology. The original books were notable for their absence of illustration, perhaps the first books on geology for 150 years to have chosen to allow the words alone to paint the scene. The inclusion here of some simple, but almost entirely uninformative, relief maps breaks this convention.

These ambitions are mistaken. The novelty and freshness of the original project was precisely because it wandered, serially, anecdotally, across America and geology, dipping in and out of subjects, of processes, theories and ages of geological time, as each new stone was overturned to set off some new story or reflection. As a commentator McPhee is excellent, but he lacks the scientific perspective to gauge a debate accurately or to check sources. He describes, for example, Scotland as laden with two miles of ice in the last Ice Age — a threefold exaggeration. It is also a pity that the one personality missing from the story is McPhee himself, for which, to find out about some of the other anthro-

pological aspects of geology — the long time spent away from home, and the accompanying disruption to family life — it is worth reading his daughter Martha's fictional *Bright Angel Time*.

Yet McPhee's original project, as articulated in the first book, was actually quite profound, and his books remain very readable. Earth scientists have a debt to McPhee for his coining of the term "Deep Time", a modern restatement of James Hutton's "no vestige of a beginning, no prospect of an end", to describe the impossibility of confronting the unfathomable depths of prehistory, the inconceivable currency of millions and billions of years that becomes so glibly familiar to geologists once they have passed through the first introductory lectures. □

Robert Muir-Wood, managing editor of Terra Nova and author of *The Dark Side of the Earth* (Unwin Hyman, 1985), is at Risk Management Solutions Ltd, Peninsular House, 30-36 Monument Street, London EC3R 8LJ, UK

Monk-y puzzles

The Mathematical Tourist

by Ivars Peterson

W. H. Freeman: 1998. 256pp. \$14.95 (pbk)

Tracking the Automatic Ant

by David Gale

Springer: 1998. 241pp. \$30, £22.50

Andrew Bremner

The ant is a fascinating object of study, and anyone who has hefted the compendious, weighty (and Pulitzer prize winning) tome *The Ants* by Bert Hölldobler and Edward O. Wilson, surely has tremendous regard for these creatures. Mathematicians, too, have now taken the ant to their hearts, and it provides one of the common links between the two books under review, which are compilations of material that has previously appeared in *Science News* and *The Mathematical Intelligencer* respectively. The former book is an updated version of Peterson's very successful first edition, published almost 10 years ago; the second collects the "Mathematical Entertainment" columns that appeared in the *Intelligencer* between 1991 and 1996, while Gale was editor.

The books differ in their emphasis. Peterson sets out to explain to the general reader (one assumes that the general reader of *Science News* has a modest mathematical background) some of the startling discoveries and recent advances that have caused excitement in mathematics. This is a difficult task, because it requires the author to have the ability not only to grasp thoroughly the mathematical concepts involved, but also to translate these concepts into entertaining prose for the non-specialist reader.

Mathematicians are really monks, but rare indeed is the illuminating manuscript. It

is thus a pleasure to find a journalist like Peterson who can write so lucidly, on rare occasions perhaps a trifle earnestly, about knots and DNA, about crystallography and Penrose tilings, about discoveries shedding light on the relationship between four-dimensional geometry and the physical theories of the nature of time, and about cryptography and prime factorization.

Occasionally, the author breathlessly takes on Baedeker: "beauty spots", "unnamed wonders", and "innumerable strange sights to view" can be found if the tourist in the Mandelbrot Set explores the area with real part between 0.26 and 0.27 and imaginary part between 0 and 0.01i ("or why not try -0.76 to -0.74 and 0.01i to 0.03i?"). *The Mathematical Tourist* conveys vividly the excitement, the usefulness, and the sheer beauty of the subject.

Gale, in contrast, writes for the pleasure-seeking mathematician, professional and amateur. These columns have a similar form to those of Martin Gardner and Ian Stewart in *Scientific American*, with an emphasis on recreational mathematics, though as we know, this is often underpinned by eminently serious stuff. Extracting tidbits for mention here is difficult to do fairly; the book contains as much to savour on each page as a compendium of short stories by Saki.

Try for instance the following problem. Each of two boxes contains an integer printed on a card; the two integers are known to be distinct. You open one of the boxes at random, and then have to guess whether the other contains a larger or a smaller integer. Is there anything you can do to give yourself a better than even chance of guessing correctly? The surprising answer is yes. Or what about the (simple) theorem that a sufficiently high power of two will have leading digits that encode the works of Shakespeare? "To be ..." for example appears at the start of $2^{9965483} = 20150205...$ A five-line algorithm is provided for computing the first such power to yield a chosen sequence, which raises interesting possibilities: perhaps a unique personalized present for that in-law who has everything (provided of course your PC can cope).

What becomes apparent from reading these books is the role now played by computers in mathematics, suggesting a shift in nature of the subject from deductive science to empirical science. For the technophobe this may appear discouraging, perhaps even alarming, with the discovery that machines have now been used to prove theorems. But the ability to amass vast amounts of experimental data certainly has enormous ramifications; without it much of the territory explored in these books would remain uncharted. □

Andrew Bremner is in the Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804, USA.

Simulated influence of carbon dioxide, orbital forcing and ice sheets on the climate of the Last Glacial Maximum

Andrew J. Weaver, Michael Eby, Augustus F. Fanning & Edward C. Wiebe

School of Earth and Ocean Sciences, University of Victoria, PO Box 3055, Victoria, British Columbia, V8W 3P6, Canada

A coupled atmosphere–ocean–sea-ice model is used to investigate the climate of the Last Glacial Maximum (~21,000 years ago) and the relative climate-forcing effects of atmosphere CO₂, the Earth's orbital parameters and ice-sheet albedo. Tropical temperatures are found to be ~2.2 °C less than today's—slightly colder than indicated by the CLIMAP palaeoclimate reconstruction. This result is consistent with a low to medium climate sensitivity to radiative perturbations. Temperatures are colder still in the northern North Atlantic region, owing to a weakening and shallowing of the thermohaline circulation. A sensitivity analysis suggests that changes in ocean circulation since the Last Glacial Maximum have not contributed directly to the global-mean temperature change since that time.

Early CLIMAP^{1,2} attempts to reconstruct sea surface temperature (SST) for the Last Glacial Maximum (LGM), around 21 kyr ago, have suggested that relative to the present, global SSTs were on average 1.7 °C cooler in August and 1.4 °C cooler in February. These reconstructions further suggested that tropical SSTs were similar to those of the present climate whereas in the North Atlantic, SSTs were much colder. Although recent alkenone evidence^{3–6} also supports tropical SSTs having been only slightly cooler at the LGM, additional evidence is contradictory. For example, coral records from Barbados⁷ and the southwest Pacific⁸, ice-core records from Peru⁹, noble-gas measurements in Brazil¹⁰ and ocean core records from the western equatorial Atlantic¹¹ suggest that LGM tropical temperatures were significantly below the present-day values.

Most previous modelling studies of the LGM have fallen into two categories. In the first, ocean-only models have been used in which surface forcing consists of restoring temperature and salinity to LGM reconstructions with specified atmospheric general circulation model (AGCM) LGM surface wind stress fields^{12–14}. The second class of studies involves the integration of AGCMs with either fixed SST or mixed-layer ocean models at the lower boundary^{15–19}. Many of these modelling studies have found tropical SSTs consistent with CLIMAP, although this often (but not always; see, for example, ref. 15) reflects the fact that CLIMAP data were used as a boundary condition. Webb *et al.*²⁰ on the other hand obtained relatively cool tropical SSTs using an AGCM in which present-day oceanic heat transports were maintained. They argued that reduced evaporation in the subtropics with present-day oceanic heat transport led to reduced water vapour in the atmosphere, and hence reduced absorption of outgoing longwave radiation. These workers also suggested that this cooling mechanism was amplified by other processes in the system.

Until recently, it has not been possible to run coupled atmosphere–ocean GCM experiments to equilibrium for LGM scenarios owing to the necessity of applying explicit flux adjustments to keep stable the present-day climate. Any meaningful comparison between the simulated present-day and LGM climates would require the use of the same flux adjustment in the LGM case. The use of flux adjustments is only valid for small perturbations away from the present climate, whereas the LGM clearly represents a large perturbation. A recent study²¹ did, however, use a coupled GCM

under LGM forcing, albeit for a short (15-year) integration by which time the deep ocean was still far from equilibrium. Here we use a coupled model of intermediate complexity, which does not employ explicit flux adjustments to the heat and freshwater air–sea flux fields. The ocean component of the coupled model is a three-dimensional GCM, whereas the atmospheric component is an energy–moisture balance model (EMBM). This analysis takes a different, yet complementary, approach to another recent modelling study²² where a zonally averaged ocean model was coupled to a simple dynamical atmospheric model with parametrized synoptic variability. Our focus will be on changes in the three-dimensional ocean circulation and its feedback on LGM climate. We will show that changes in ocean circulation do not contribute to the global LGM cooling, and that the resulting tropical SSTs are slightly colder than CLIMAP, consistent with alkenone data.

The model

The atmospheric component of our coupled model²³ consists of energy and moisture balance equations for heat and fresh water. Poleward heat and freshwater transport are parametrized through Fickian diffusion, and precipitation is assumed to occur when the relative humidity reaches greater than 85%. Although the EMBM does not retain any explicit dynamics, a wind feedback parametrization is employed. During the spin-up to equilibrium, monthly climatological wind-stress fields were used at the ocean surface²⁴. Wind speeds were also derived from these fields²³ for use in the latent and sensible heat flux parametrizations at the air–sea interface. In the LGM experiments, perturbations of annual mean sea-level pressure away from the present-day equilibrium allowed for the calculation of anomalous geostrophic (frictional near the Equator) winds which were then rotated and contracted to yield wind stress/speeds. These in turn were added to the present-day fields²⁵. This approach therefore leads to a first-order approximation of atmospheric circulation feedbacks in response to changing surface air temperature (SAT) and humidity and hence sea-level pressure fields.

A thermodynamic sea-ice model²⁶ and a parametrization of water vapour/planetary longwave feedbacks²⁷ were also included in the model. Ice albedo feedbacks were incorporated, whenever sea ice was present and whenever the air temperature over land was below

-10°C , by locally reducing the latitudinal profile of the co-albedo by 0.1 (ref. 28). In the LGM experiment, we also used this local co-albedo reduction at all land points where ice sheets were thought to be present²⁹. The ocean component of the coupled model is based on the GFDL model³⁰ as discussed in ref. 31. Precipitation over land was assumed to instantaneously return to the ocean via one of 21 river drainage basins (Fig. 1a). The coupled model has 3.75° zonal, and 1.8555° meridional, resolution with 19 vertical levels in the ocean.

The only differences between the LGM experiments and the present-day experiments are imposed changes in orbital parameters, atmospheric CO_2 concentrations and the presence of ice sheets (via an albedo feedback). The orbital parameters were set to values appropriate³² for 21 kyr ago and an atmospheric CO_2 concentration of 200 p.p.m. (compared with 350 p.p.m. for the present climate) was used. This CO_2 concentration was converted to a radiative forcing according to ref. 33 whereby a doubling of CO_2

leads to a radiative forcing of 4 W m^{-2} , translating to a radiative forcing of -3.2 W m^{-2} for the LGM. In our model, the equilibrium response for CO_2 doubling is a 3°C increase in global-mean SAT so that its climate sensitivity is $0.75^{\circ}\text{C W}^{-1}\text{ m}^2$.

Simulations of the LGM and the present day

Four sets of simulations (experiments) were performed, with each set differing only in the river mask (drainage basins) that was used in the region around the North Atlantic. The reason for this sensitivity study will be apparent later. At this stage we address the results from the first suite of experiments (model 1) with a river mask (Fig. 1a) that leads to the most realistic representation of the present-day thermohaline circulation.

The first experiment in the model 1 suite was an integration of the model to equilibrium under present-day orbital and CO_2 forcing. In the North Atlantic, about 16 Sv of deep water formed at high latitudes with about 8 Sv being exported to the Southern Ocean

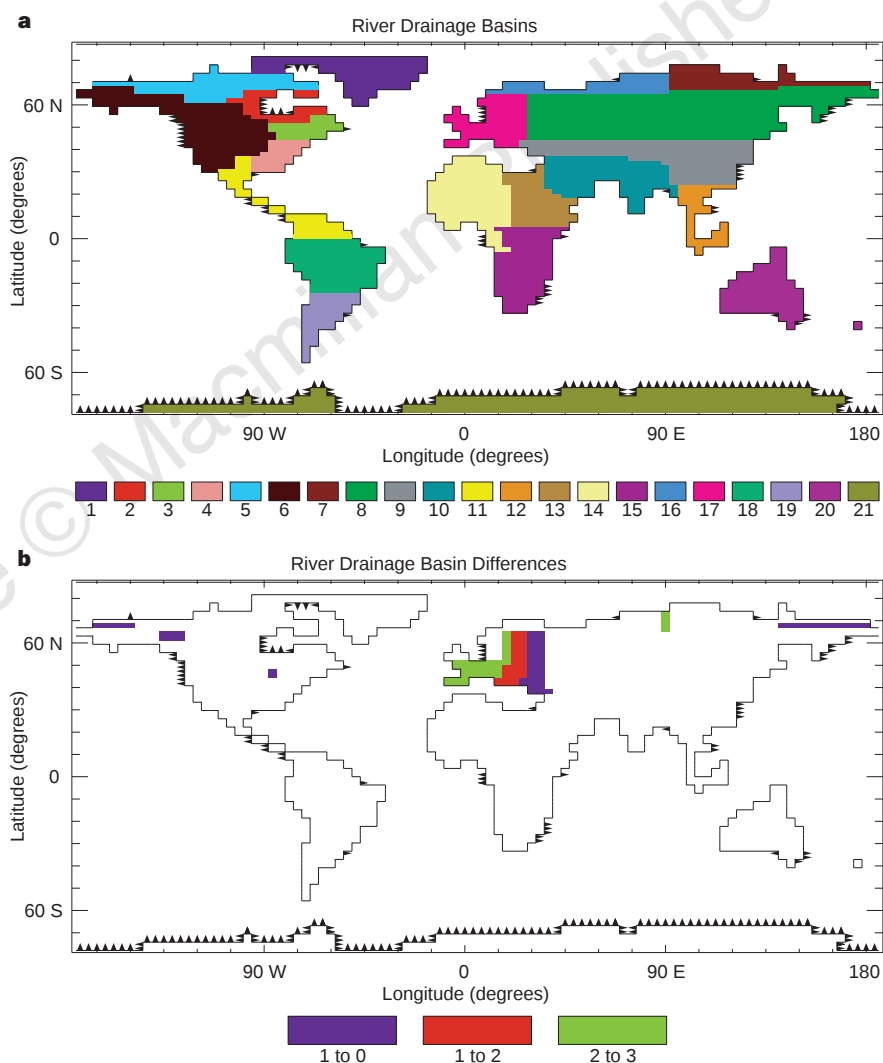


Figure 1 River drainage basins. **a**, Drainage basins for the 21 different areas defined. The numbers in the colour key refer to the following river basins: (1), Greenland; (2), Hudson Bay; (3), St Lawrence; (4), Mississippi; (5), Mackenzie; (6), Columbia; (7), Lena; (8), Amur; (9), Yangtze-Yellow; (10), Ganges; (11), Panama; (12), Mekong; (13), Nile; (14), Congo; (15), Lumpapa-Zambezi; (16), Yenesei-Ob; (17), Rhine; (18), Amazon; (19), La Plata; (20), Murray; (21), Antarctica. **b**, Changes made to the drainage basins between model 1 (the control) and models 0, 2 and 3. The dark blue area indicates changes between model 1 and model 0. These involved an expansion of the Rhine catchment area with reductions to both the Yangtze-

Yellow and Amur basins. In addition, small diversions were made from the Amur to the Lena, the Columbia to the St Lawrence, and the Columbia to the Mackenzie. The red indicates the changes between model 2 and model 1 which involved a reduction in the Rhine catchment area via an expansion of the Yangtze-Yellow and Amur basins. The green indicates further changes from model 2 to model 3. Here the Rhine catchment area is still further reduced with the Amur basin expanding more broadly westward. The arrows in **a** and **b** indicate discharge points for the particular drainage basin.

above 3,000 m depth (Fig. 2a). The deep Atlantic, south of about 30° N, is filled with Antarctic Bottom Water (AABW) formed in the Weddell and Ross seas. In the second experiment, the atmospheric CO₂ level was set to 200 p.p.m. and orbital parameters for 21 kyr ago were used. In addition, an ice-sheet albedo feedback was included wherever ice sheets were present in the reconstruction of ref. 29. The first important finding from our analysis is the weakening and shallowing, but not collapse, of the conveyor in the North Atlantic, and the enhanced intrusion there of AABW (Fig. 2b). The actual location of North Atlantic Deep Water (NADW) formation did not change and North Pacific Intermediate Water (NPIW) did not intensify, both in contrast with the results of ref. 22. In fact, we found a 50% reduction in NPIW formation (from 2.2 Sv to 1.1 Sv)

in our LGM climate. This discrepancy may lie in the fact that a zonally averaged ocean model with coarse meridional resolution (10°) was used in ref. 22. It is unlikely that the processes involved in the formation of NPIW (overflow from the Sea of Okhotsk and subsequent mixing of Oyashio waters with overlying Kuroshio waters in the Kuroshio extension) are well resolved in either of our models.

Associated with this thermohaline weakening and shallowing is a reduction in the heat transport in the North Atlantic (Fig. 2c). The results shown in Fig. 2c are in contrast with those in ref. 20 where it was assumed that ocean heat transport did not change in the LGM. The Atlantic heat transport curve in Fig. 2c has reduced everywhere in the North Atlantic (and subsequently the Northern Hemisphere). This was compensated for by an increase in the heat transport in the Southern Hemisphere.

The SST difference (Fig. 3a) is in excellent agreement with alkenone reconstructions^{3–6,34–36}. In particular, our tropical SST results are colder than those of CLIMAP^{1,2}, consistent with alkenone reconstructions^{3–6}, but not as cold as the bore-hole, coral and other data mentioned earlier^{7–11}. Model results reveal annual and global mean SST cooling of ~2.3 °C, with ~2.0–3.4 °C in the tropical Atlantic and ~2.0 °C in the tropical Indian oceans. In the North Atlantic the signal is amplified, with up to 6.3 °C cooling associated with a weakening and shallowing of the conveyor. The Pacific Ocean shows the greatest variance across the basin, with western tropical Pacific SSTs only about 1.9 °C cooler than today and eastern tropical Pacific SSTs as much as 3.1 °C cooler than today, suggesting that LGM Pacific SSTs are similar to permanent La Niña conditions.

The SAT difference (Fig. 3b–d) shows pronounced cooling around the region of the North Atlantic due to a weakening and shallowing of the conveyor, the southward expansion of sea ice and the subsequent sea-ice albedo feedback there. Averaged over the globe, the LGM experiment cooling is 3.2 °C, with more intense cooling in the Northern Hemisphere (where there is also more land) than in the Southern Hemisphere, yielding a pronounced asymmetry about the Equator. Cooling of 3.2 °C is slightly less than found in some AGCM studies^{16–18}, although it is similar to the amount found in others^{37,38}. In the Northern Hemisphere maximum winter-mean cooling of 14.4 °C is found in the North Atlantic; this weakens to about 6.8 °C in the summer. In the Southern Hemisphere, the seasonal-mean variation in maximum cooling over Antarctica (near the Ross Sea) ranges from 12.9 °C in winter to 5.7 °C in summer (Fig. 3c, d). The sea-ice distribution in the LGM experiment (solid line in Fig. 3c–d) bears reasonable agreement with various reconstructions^{39,40}. Of particular importance is the result that most of the North Atlantic is ice free when averaged over the summer months⁴¹.

To facilitate a detailed comparison between observed (CLIMAP^{1,2}) and model SST estimates, we first subtract the CLIMAP present-day reconstruction from the CLIMAP LGM reconstruction, for the month of February (winter) and again for the month of August (summer). We then compute the SST difference between our modelled LGM and present-day experiments for February and then for August. Finally, we subtract the observational difference field from our model difference field. In both the summer (Fig. 4a) and the winter (Fig. 4b), we find significantly more cooling in the subtropical North and South Pacific and Atlantic—with the exception of the subtropical North Atlantic, these are regions where CLIMAP data suggests modest LGM warming. The maximum Northern Hemisphere winter differences are 5.8 °C, 4.8 °C, 2.4 °C and 2.6 °C in the North Pacific, South Pacific, North Atlantic and South Atlantic subtropical gyres, respectively. These increase in the Northern Hemisphere summer to 6.8 °C, 3.1 °C and 3.3 °C in the South Pacific, North Atlantic and South Atlantic, respectively and decrease to 5.7 °C in the North Pacific. Our North Atlantic is also persistently less cool than CLIMAP, and our Southern Ocean has a more rich zonal structure.

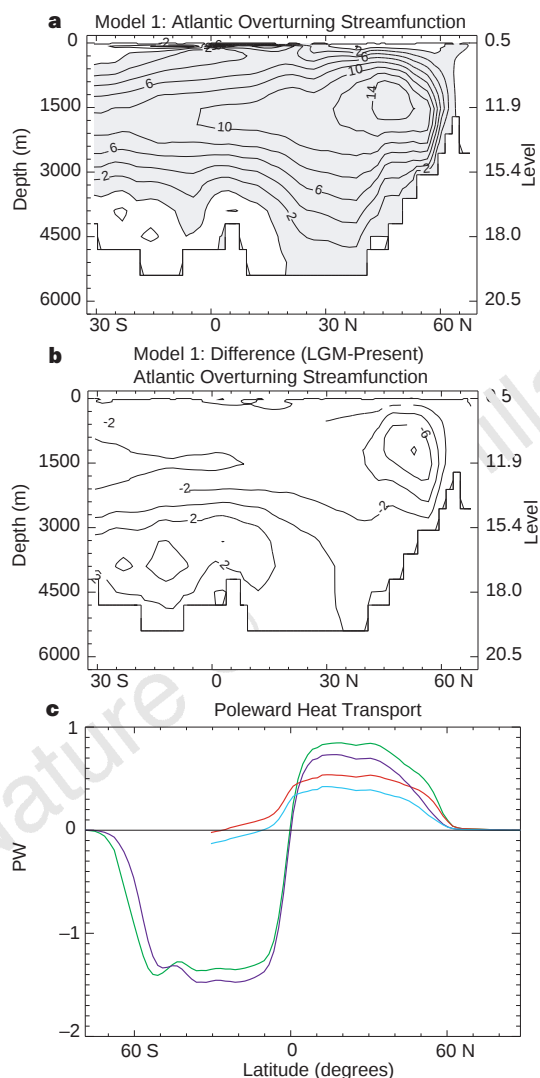


Figure 2 Present-day and LGM conveyor and poleward heat transport. **a**, Meridional overturning streamfunction in the Atlantic Ocean (contours, in Sv; 1 Sv = 10⁶ m³ s⁻¹) for the present-day equilibrium. **b**, Difference (contours, in Sv) between the LGM and the present-day meridional overturning in the Atlantic Ocean. The contour interval in **a** and **b** is 2 Sv, and clockwise circulation is shaded. **c**, Poleward heat transport in PW (1 PW = 10¹⁵ W) from the present-day and LGM experiments. The green and red lines represent the present-day global and Atlantic heat transports, respectively. The corresponding LGM curves are given by the purple and blue lines, respectively. The present-day heat transports are weaker than observational estimates, a feature which is common in coarse-resolution ocean GCMs which do not well represent narrow, intense western boundary currents. The model level corresponding to the depth is given on the right-hand y-axis.

Along the equatorial band in the Northern Hemisphere winter, our model estimates of LGM cooling differ by 3.4 °C from CLIMAP in the eastern equatorial Pacific (where CLIMAP suggests slight warming), dropping to about 1 °C in the western equatorial Pacific. The Indian and Atlantic equatorial oceans are 1–2 °C lower than CLIMAP estimates, in line with recent alkenone measurements^{3,4,6,35}. Our discrepancy with CLIMAP is less in the Northern Hemisphere summer, where our model is in fact up to 1.8 °C warmer than CLIMAP in the warm pool of the western equatorial Pacific.

The weakening and shallowing of the North Atlantic conveyor in the LGM experiment and the general surface cooling caused a redistribution of water masses in the ocean as indicated in Fig. 5. Model results suggest only a slight cooling (~0.2 °C) of AABW near the source region, with a larger (~0.5–0.7 °C) general cooling of the deep ocean, and an even larger (~1.5–3.5 °C) cooling in the thermocline waters (Fig. 5a). In the Atlantic (Fig. 5b), the cooling is pronounced in the region of NADW export to the Southern Ocean, due to a reduction of the conveyor (Fig. 1b).

Sensitivity to components of radiative forcing

We now determine the extent to which the response found in the last section is due to orbital forcing, CO₂ forcing, ice-sheet albedo feedback, or the wind-stress feedback parametrization. Four additional experiments were performed which involved repeating the LGM experiment with a particular process turned on or off. In the first of these four experiments, the ice-sheet albedo feedback was turned off. The thermohaline circulation still weakened under the combined effects of orbital and atmospheric CO₂ forcing, although the global mean SAT and SST were now only 2.5 °C and 1.8 °C cooler than the present, respectively. This shows that the addition of the albedo feedback accounted for about 0.7 °C and 0.5 °C of the global mean SAT and SST cooling, respectively, although as discussed later, the role of ice-sheet albedo feedbacks may be underestimated in our model. The spatial pattern of the difference clearly showed intensified cooling over the locations of the Laurentide, Fennoscandian and Antarctic ice sheets which in turn caused diffusive heat transport from the oceans to land and subsequent SST cooling.

We now use the results of the experiment without the ice-sheet

albedo feedback as the reference to examine the sensitivity to other feedbacks. The removal of the wind stress feedback had little effect on the climate response, although there were small local changes in the North Atlantic where the feedback was strongest. When the orbital forcing was turned off (by setting orbital parameters to present-day values) the resultant annual-mean response of the model was almost identical to the response with it included, although the seasonal variation in SAT was increased. Global-mean SAT and SST were now 2.5 °C and 1.9 °C cooler than the present, respectively. These numbers are virtually the same as those reported with no ice-sheet albedo feedback.

When the atmospheric CO₂ forcing was turned off, leaving the orbital parameters at values for 21 kyr ago, there was only a slight cooling from the present day in the annual and global mean SAT (0.4 °C) and SST (0.1 °C). This slight cooling was caused by an enhanced ice albedo feedback which existed in high latitudes, as more sea-ice was present in the summer due to a decreased seasonal variation in SAT. Nevertheless, the reason why the previous experiment (with no orbital forcing) captures the global mean SST and SAT cooling response found when both orbital and CO₂ forcing are included, is that the small changes in sea-ice extent due to the reduced seasonal cycle are encompassed within the larger changes in sea-ice extent due to the changes in atmospheric CO₂.

Sensitivity to initial present-day conveyor strength

Two studies^{42,43} have demonstrated the potential instability of the North Atlantic conveyor in ocean GCMs when the overturning rate is near a stability threshold. We examine this in our model by conducting a series of experiments, identical to those described in the previous two sections, but now with slightly different initial present-day conveyor strengths. It is well known⁴⁴ that the strength of the conveyor is very sensitive to the amount of freshwater discharge into the North Atlantic. In addition, a significant uncertainty in our model concerns what river drainage basins should be imposed, as we have no topography on land and hence precipitation is not induced by topographical features such as mountain ranges. Therefore, to accomplish the task of creating several initial strengths of the conveyor, we modified the catchment areas of a few rivers

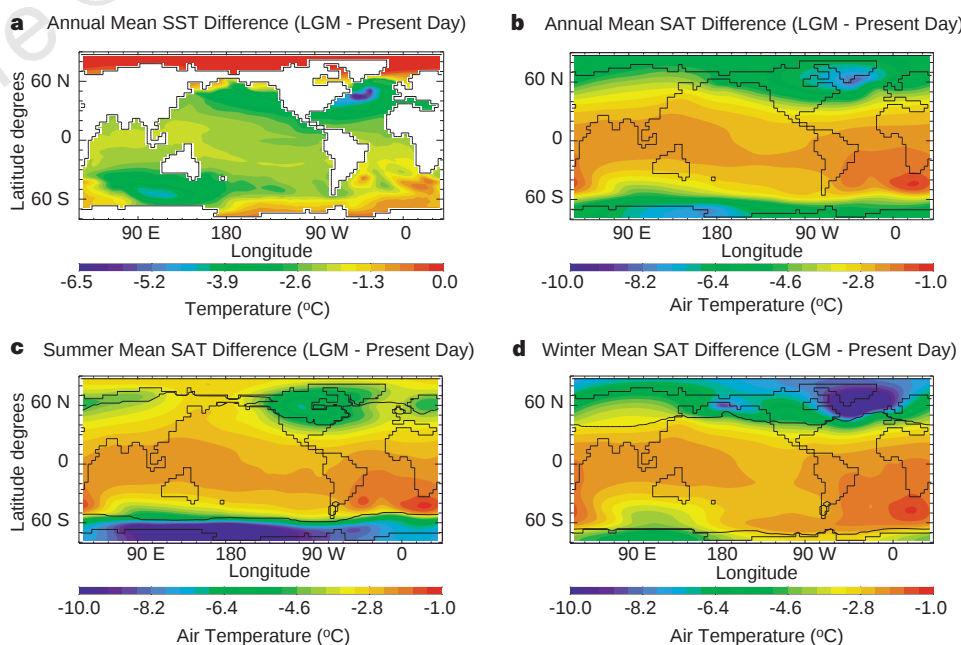


Figure 3 Difference fields between the LGM and present-day experiments. **a**, Annual mean SST; **b**, annual mean SAT; **c**, Northern Hemisphere summer SAT; **d**, Northern Hemisphere winter SAT. The black line over the ocean in **c** and **d** indicates the maximum sea-ice extent whereas on land it indicates the latitude of

the –10 °C contour or the ice-sheet boundary (whichever extends further equatorward). By showing sea-ice extent averaged over a season, the contour will show the maximum sea-ice extent during that time. The minimum extent in summer is further north than indicated in **c**.

around the North Atlantic, thereby changing the amount of fresh-water discharge into the ocean there.

The modifications that were done are illustrated in Fig. 1b. In model 0, the main change was an increase of the Rhine catchment area (over model 1) at the expense of the Amur and Yangtze-Yellow rivers. This had the effect of diverting more fresh water into the North Atlantic instead of the Pacific, with the result that the conveyor had a strength of 12 Sv at the present-day equilibrium (Fig. 6a). When LGM forcing was used with the model 0 drainage basins, a collapse of the conveyor (Fig. 6b) and hence North Atlantic heat transport (Fig. 6c) ensued. In model 2, the Rhine catchment area was decreased slightly from that of Fig. 1a by increasing the discharge into the Amur and Yangtze-Yellow rivers. The net result was a slightly stronger conveyor (now 17 Sv; Fig. 6d) although AABW was beginning to disappear in the deep North Atlantic. The changes made in model 3 involved a more drastic further reduction of the Rhine and increase of the Amur basins. This led to a further increase in the present-day conveyor (to about 18 Sv) but the near elimination of deep AABW in the North Atlantic (Fig. 6g). The choice of the model 1 mask over the other three masks is now clear: model 0 yields too weak an NADW signal whereas models 2 and 3 yield too weak an AABW signal. Model 1, on the other hand, gives an NADW strength of 16 Sv, in good agreement with observations⁴⁵, yet not too strong to prevent the intrusion of AABW.

When forcing characteristic of 21 kyr ago was used in both the model 2 and 3 suite of experiments, the results were similar to those of model 1, although differences existed. In particular, the stronger the present-day conveyor was at equilibrium (Figs 2a, 6a, g), the further it was from a stability threshold and so the less sensitive it was to 21-kyr-ago forcing (Figs 2b, 6e, h). In fact, the equilibrium heat transport for the LGM changed less for a stronger present-day conveyor (Figs 2c, 6f, i). This result is more in line with the

assumption used in ref. 20 although convergence was not reached and the present-day thermohaline circulation became less realistic. Indeed in model 3, the Atlantic and hence global heat transport in fact increased under LGM forcing. For all our present-day climatologies, the discussion of CO₂, orbital forcing and ice-sheet feedbacks given in the last section remained valid.

Possibly the most remarkable result is that for all four LGM experiments, starting from four different realizations of the present-day ocean climate, the global-mean SST and SAT cooling at equilibrium were virtually the same (~ 2.3 and ~ 3.2 °C, respectively), although significant regional differences existed. Tropical SST differences averaged over the latitude band 20° S to 20° N were also the same in these four experiments (2.2 °C cooling). This suggests that redistributions in ocean circulation did not cause enhanced global mean cooling in the LGM unless they did so indirectly, via changes in CO₂ release/uptake. This is in contrast with results in ref. 22 where a 30% contribution to global cooling was found, due to a reorganization of the ocean circulation. This discrepancy may be due to the more southward LGM sea-ice extent in ref. 22 (and the subsequent ice-albedo feedback) due to a southward shift in the location of NADW formation. In addition, the lack of an ocean feedback contributing to enhanced global cooling accounts for most of the difference in LGM cooling between results of ref. 22 and our results.

Interpretations and implications

Although our EMBM is admittedly idealized, the present study represents an attempt to examine quantitatively the climate of the

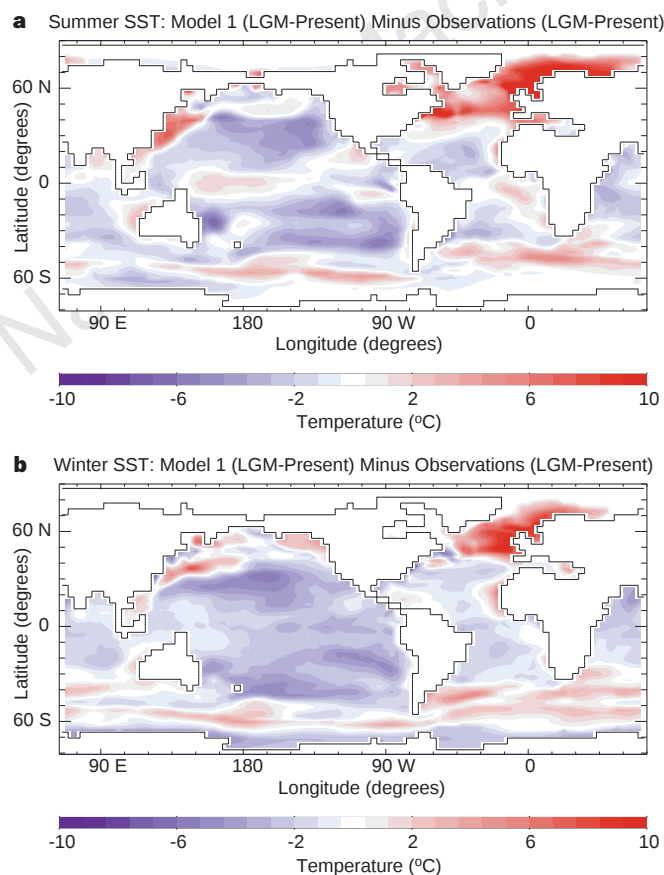


Figure 4 Anomaly maps showing the difference between model and CLIMAP¹² SST changes between the LGM and the present day. **a**, Summer; **b**, winter.

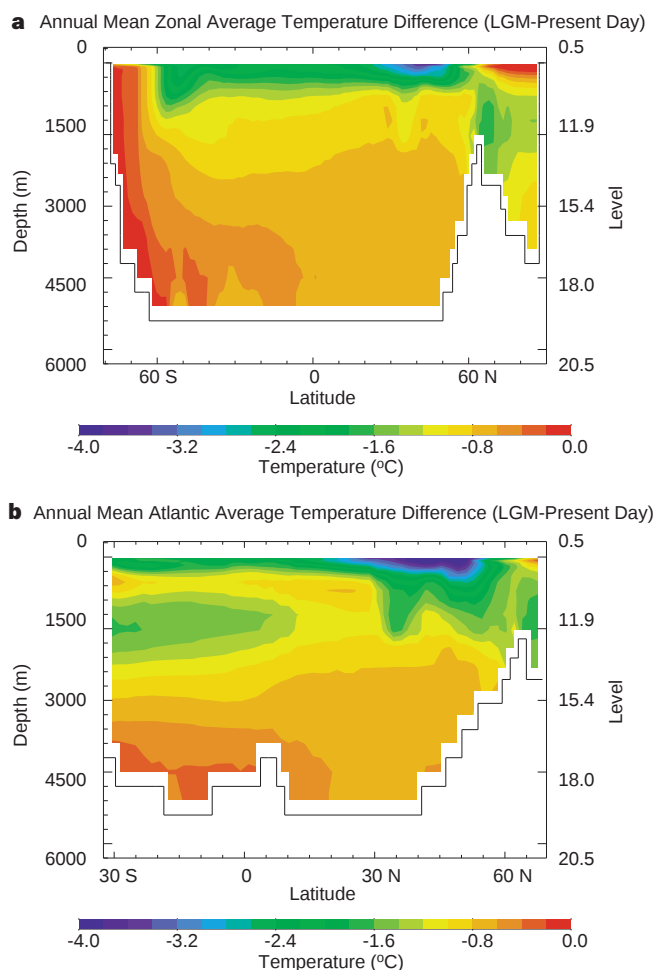


Figure 5 Zonally averaged ocean potential temperature difference fields between the LGM and present-day experiments. **a**, Global average; **b**, Atlantic average. The model level corresponding to the depth is given on the right-hand y-axis.

LGM and its sensitivity to ice-sheet albedo, orbital forcing and atmospheric CO₂ using a coupled model which is integrated to equilibrium. Our experiments were purposely kept simple in order to try and reveal the competing effects of the different forcing mechanisms. We have not included the radiative forcing associated with atmospheric aerosols for a number of reasons. First, it is not clear what the spatial pattern of this aerosol forcing would be in terms of both aerosol type and distribution. For example, estimates from Peru⁹ suggest that LGM atmospheric dust levels may have been 200 times the present, whereas in Antarctica they may have been 4–6 times the present for marine and 10–30 times the present for continental aerosols⁴⁶. In addition, it is not clear whether these aerosols would contribute to local cooling or warming through the competing effects of increased albedo and longwave absorption.

The comparison between model results and proxy data over land is complicated by the fact that there is no topography in our EMBM. We would therefore expect, through purely adiabatic cooling, our results over land to be warmer than those in the proxy record at higher elevations. This follows since latent-heat release in a warmer climate would tend to yield a generally smaller globally averaged adiabatic lapse rate than in a colder climate, due to the nonlinearity of the Clausius–Clapyeron equation. In addition, vegetation feedbacks and subsequent albedo changes through the expansion of dry-land vegetation in Australia and Africa, and the conversion of conifer to tundra in Europe and Siberia, would enhance the cooling there¹⁹. Other potentially important factors were not included in this model; we used the present-day land boundaries which did not account for the sea level drop of 121 ± 5 m (ref. 47) (and hence global mean salinity increase) in the LGM, and the atmospheric model did not allow for changes in cloud cover. We purposely retained the present-day land boundaries in the LGM experiments

to facilitate comparison with model results from the present-day experiment, although additional experiments not presented here with a 120-m sea-level drop accounted for, and with the global mean salinity increased by 1 p.s.u., revealed virtually no differences. In addition, the retention of present-day land boundaries makes easier the understanding of the competing effects of orbital, atmospheric CO₂, and ice-sheet albedo effects. Reduced LGM cloud cover¹⁶ in a drier atmosphere probably contributed to extra cooling, although the incorporation of this effect in our vertically averaged EMBM would be difficult, and perhaps inconsistent with its level of sophistication. In addition, the lack of topography on land does not allow us to capture the topographical effects of ice sheets, which have been shown to act as a summer heat sink (through local melting) and to influence the atmospheric circulation causing the advection of very cold air across the North Atlantic¹⁵. Our reduction of the planetary co-albedo by only 0.1 is also probably an underestimate for the ice albedo. This may explain why CO₂ is a more effective mechanism for cooling in our model, in distinction to the results of ref. 37 where it is of equal importance with ice-sheet effects.

With these shortcomings recognized, a number of conclusions can be drawn from this study. First, the assumption of invariant ocean heat transports between the LGM and the present²⁰ is not consistent with our findings. Second, in response to LGM radiative forcing we found a weakening and shallowing (but not collapse) of the North Atlantic conveyor, consistent with observations^{48,49}. Our model SSTs are in good agreement with alkenone reconstructions^{3–6,34–36}, and are generally cooler than those of CLIMAP^{1,2} in the tropical ocean, while in the North Atlantic, our model results suggest a maximum SST cooling of $\sim 6.3^\circ\text{C}$ with most of the cooling in the 4° – 6°C range (noticeably lower than CLIMAP^{1,2}

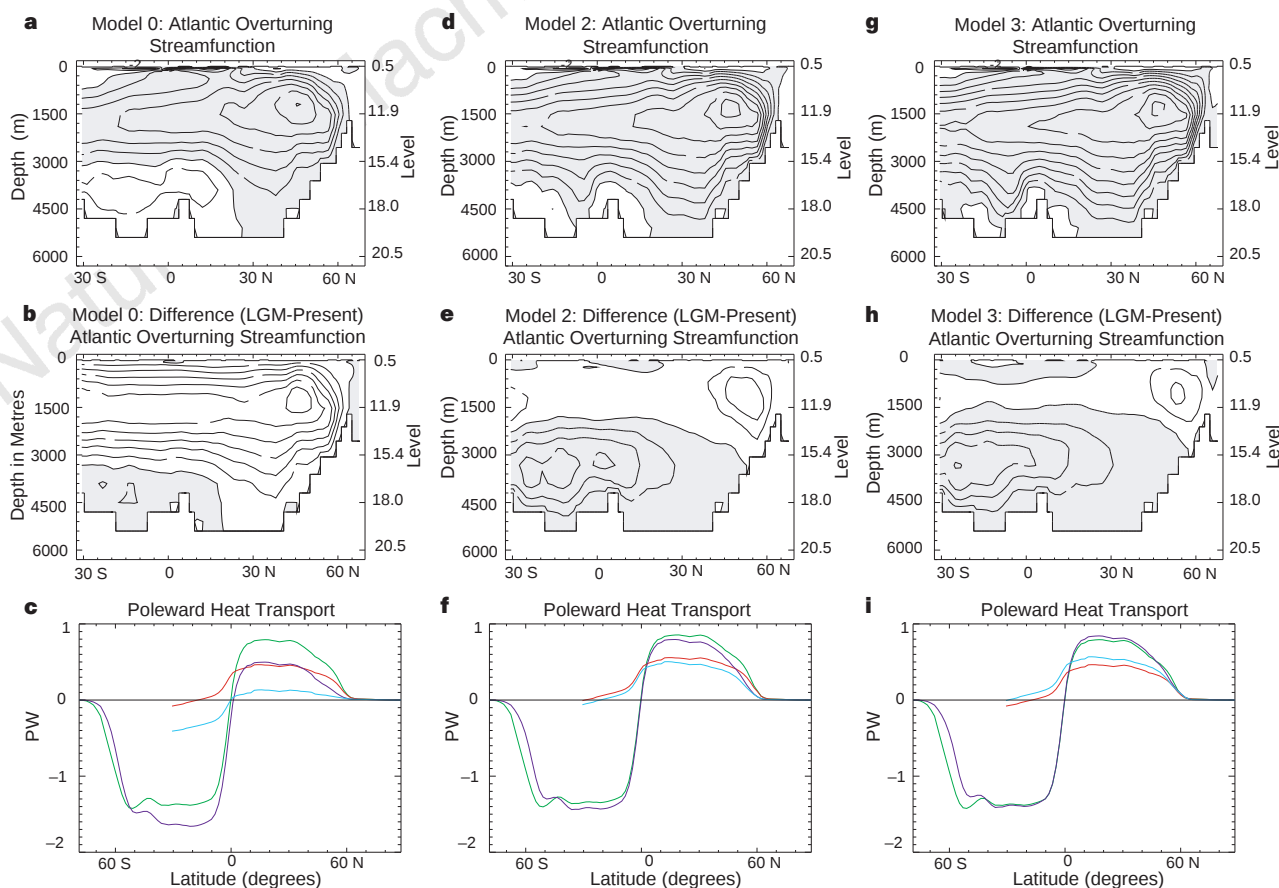


Figure 6 As in Fig. 2 but for the three other experiments with different initial present-day conveyor strengths in the North Atlantic. The drainage basins used to derive the different initial conditions, for the purpose of the sensitivity analysis,

are described in Fig. 1. Model 0 (a–c) gives a weak initial conveyor (12 Sv); model 2 (d–f) and model 3 (g–i) yield stronger initial conveyor strengths than shown in Fig. 2a (17 Sv and 18 Sv, respectively).

estimates). Some of this reduced cooling may be due to our present-day climatology already being slightly colder than observations in the Nordic seas, a problem that is common in coarse-resolution ocean GCMs. Large differences from CLIMAP also exist in the subtropical gyres of all oceans.

For all our LGM experiments starting from different present-day initial conditions, the global-mean SST and SAT cooling at equilibrium were virtually the same, although significant regional differences existed. This result is important, in that it suggests that redistributions in ocean circulation did not directly cause global mean cooling in the LGM, although indirect effects via changes in CO₂ release/uptake are still plausible. The examination of the relative importance of orbital forcing, atmospheric CO₂ levels and ice-sheet albedo feedbacks suggests that the most important of these forcings in our model is the change in atmospheric CO₂, with ice-sheet albedo feedbacks contributing to further cooling. Nevertheless, changes in orbital forcing, through decreased seasonality, do induce slight expansions in sea-ice extent.

Three studies have now been conducted with dynamical ocean models coupled to atmospheric models of various degrees of complexity. In the first of these²², a coarse resolution, zonally averaged ocean model was coupled to a simple dynamical atmospheric model and integrated to equilibrium; in the second²¹, ocean and atmospheric GCMs were coupled together and integrated under LGM forcing for 15 years (starting from a present-day initial condition). The present study (the third) involves the integration to equilibrium of an ocean GCM coupled to an idealized EBM. The climate sensitivity of our model is similar to the model used in ref. 22 (3.0 °C warming upon CO₂ doubling), whereas the GFDL model has a slightly higher sensitivity (3.7 °C)⁵⁰. All three of these studies find tropical temperatures which are colder than CLIMAP^{1,2}, although intermodel differences exist due to the magnitude of the ocean feedback and the sensitivity of each model to its radiative forcing. Our results with low-to-moderate climate sensitivity and no ocean feedback to LGM cooling yield SSTs which are slightly colder than CLIMAP. Ganopolski *et al.*²², once more with a low-to-moderate climate sensitivity, predict cooler tropical temperatures largely through a positive feedback from changes in ocean circulation. Bush and Philander²¹, with a moderate-to-high climate sensitivity, predict cooler tropical temperatures (with global LGM cooling of only 4.3 °C, compared to 3.2 °C here and 6.2 °C in ref. 22). It is possible for us to reconcile a more stable tropics (colder than CLIMAP^{1,2}, consistent with alkenone reconstructions^{3–6}, but not as cold as the bore-hole, coral and other reconstructions^{7–11} mentioned earlier) if the climate sensitivity is low-to-moderate and there is little or no thermohaline feedback. Nevertheless, before such a conclusion is firmly established more effort needs to be directed to understand the following: first, whether or not changes in the ocean circulation can cause a positive feedback on the global-mean LGM cooling via changes in deep and intermediate water production²² or extratropical subduction²¹; second, what the sensitivity is of the real climate system, and how the parametrization of the physical processes of cloud and water-vapour feedbacks affects this climate sensitivity. □

Received 5 December 1997; accepted 29 May 1998.

- CLIMAP Project Members The surface of the ice-age earth. *Science* **191**, 1131–1137 (1976).
- CLIMAP Project Members Seasonal reconstructions of the earth's surface at the glacial maximum. (Map Chart Ser. MC-36, Geol. Soc. Am., Boulder, CO, 1981).
- Sikes, E. L. & Keigwin, L. Equatorial Atlantic sea surface temperature for the last 30 kyr: A comparison of U₁₇, δ¹⁸O and foraminiferal assemblage temperature estimates. *Paleoceanography* **9**, 31–45 (1994).
- Bard, E., Rostek, F. & Sonzogni, C. Interhemispheric synchrony of the last deglaciation inferred from alkenone palaeothermometry. *Nature* **385**, 707–710 (1997).
- Lyle, M. W., Prah, F. G. & Sparrow, M. A. Upwelling and productivity changes inferred from a temperature record in the central equatorial Pacific. *Nature* **355**, 812–815 (1992).
- Rostek, F. *et al.* Reconstructing sea surface temperature and salinity using δ¹⁸O and alkenone records. *Nature* **364**, 319–321 (1993).
- Guilderson, T. P., Fairbanks, R. G. & Rubenstein, J. L. Tropical temperature variations since 20,000 years ago: modulating interhemispheric climate change. *Science* **263**, 663–665 (1994).
- Beck, J. W. *et al.* Abrupt changes in early Holocene tropical sea surface temperature derived from coral records. *Nature* **385**, 705–707 (1997).

- Thompson, L. G. *et al.* Late glacial stage and Holocene tropical ice core records from Huacarán, Peru. *Science* **269**, 46–50 (1995).
- Stute, M. *et al.* Cooling of tropical Brazil (5 °C) during the last glacial maximum. *Science* **269**, 379–383 (1995).
- Curry, W. B. & Oppo, D. W. Synchronous high-frequency oscillations in tropical sea surface temperatures and North Atlantic deep water production during the last glacial cycles. *Paleoceanography* **12**, 1–14 (1997).
- Lautenschlager, M., Mikolajewicz, U., Maier-Reimer, E. & Heinze, C. Application of ocean models for the interpretation of atmospheric general circulation model experiments on the climate of the Last Glacial maximum. *Paleoceanography* **7**, 769–782 (1992).
- Fichefet, T., Hovine, S. & Duplessy, J.-C. A model study of the Atlantic thermohaline circulation during the last glacial maximum. *Nature* **372**, 252–255 (1994).
- Winguth, A. M. E. *et al.* On the Sensitivity of an Ocean General Circulation Model to Glacial Boundary Conditions (Rep. No. 203, Max-Planck-Institut für Meteorologie, Hamburg, 1996).
- Manabe, S. & Broccoli, A. J. The influence of continental ice sheets on the climate of an ice age. *J. Geophys. Res.* **90**, 2167–2190 (1985).
- Hansen, J. *et al.* in *Climate Processes and Climate Sensitivity* 130–163 (Geophys. Monogr. 29, Am. Geophys. Union, Washington DC, 1984).
- Lautenschlager, M. & Herterich, K. Atmospheric response to ice age conditions: climatology near the earth's surface. *J. Geophys. Res.* **95**, 22547–22557 (1990).
- Kutzbach, J. E. & Guetter, P. J. The influence of changing orbital parameters and surface boundary conditions on climate simulations for the past 18000 years. *J. Atmos. Sci.* **43**, 1726–1759 (1986).
- Crowley, T. J. & Baum, S. K. Effect of vegetation on an ice-age climate model simulation. *J. Geophys. Res.* **102**, 16463–16480 (1997).
- Webb, R. S. *et al.* Influence of ocean heat transport on the climate of the Last Glacial Maximum. *Nature* **385**, 695–699 (1997).
- Bush, A. B. G. & Philander, S. G. H. The role of ocean-atmosphere interactions in tropical cooling during the Last Glacial Maximum. *Science* **279**, 1341–1344 (1998).
- Ganopolski, A., Rahmstorf, S., Petoukhov, V. & Claussen, M. Simulation of modern and glacial climates with a coupled global model of intermediate complexity. *Nature* **391**, 351–356 (1998).
- Fanning, A. F. & Weaver, A. J. An atmospheric energy-moisture balance model: climatology, interpentadal climate change, and coupling to an ocean general circulation model. *J. Geophys. Res.* **101**, 15111–15128 (1996).
- daSilva, A. M., Young, C. C. & Levitus, S. *Atlas of Surface Marine Data 1994* Vol. 3 (NOAA, Washington DC, 1994).
- Fanning, A. F. & Weaver, A. J. Temporal geographical meltwater influences on the North Atlantic Conveyor: Implications for the Younger Dryas. *Paleoceanography* **12**, 307–320 (1997).
- Semtner, A. J. A model for the thermodynamic growth of sea ice in numerical investigations of climate. *Phys. Oceanogr.* **6**, 379–389 (1976).
- Thompson, S. J. & Warren, S. G. Parameterization of outgoing infrared radiation derived from detailed radiative calculations. *J. Atmos. Sci.* **39**, 2667–2680 (1982).
- Graves, C. E., Lee, W. H. & North, G. R. New parameterizations and sensitivities for simple climate models. *J. Geophys. Res.* **98**, 5025–5036 (1993).
- Peltier, W. R. Ice age paleotopography. *Science* **265**, 195–201 (1994).
- Pacanowski, R. C. Momz user's guide and reference manual (GFDL Ocean Group Tech. Rep. No. 3, GFDL/NOAA, Princeton, NJ, 1995).
- Weaver, A. J. & Hughes, T. M. C. On the incompatibility of ocean and atmosphere models and the need for flux adjustments. *Clim. Dyn.* **12**, 141–170 (1996).
- Berger, A. L. Long term variations of daily insolation and Quaternary climatic changes. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
- Ramanathan, V. *et al.* Climate-chemical interactions and effects of changing atmospheric trace gases. *Rev. Geophys.* **25**, 1411–1482 (1987).
- Ikehara, M. *et al.* Alkenone sea surface temperature in the Southern Ocean for the last two deglaciations. *Geophys. Res. Lett.* **24**, 679–682 (1997).
- Schneider, R. R., Müller, P. J. & Ruhland, G. Late Quaternary surface circulation in the east equatorial South Atlantic: Evidence from alkenone sea surface temperatures. *Paleoceanography* **10**, 197–219 (1995).
- Chapman, M. R. *et al.* Faunal and alkenone reconstructions of subtropical North Atlantic surface hydrography and paleotemperature over last 28 kyr. *Paleoceanography* **11**, 343–357 (1996).
- Broccoli, A. J. & Manabe, S. The influence of continental ice, atmospheric CO₂, and land albedo on the climate of the last glacial maximum. *Clim. Dyn.* **1**, 87–99 (1987).
- Hyde, W. T. *et al.* Comparison of GCM and energy balance model simulations of seasonal temperature changes over the past 18,000 years. *J. Clim.* **2**, 864–887 (1989).
- Denton, G. H. & Hughes, T. J. (eds) *The Last Great Ice Sheets* (Wiley, New York, 1981).
- Cooke, D. W. & Hays, J. D. in *Antarctic Geoscience* (eds Craddock, C. *et al.*) 1017–1025 (Univ. Wisconsin Press, Madison, 1982).
- Hebbsen, D. *et al.* Moisture supply for northern ice-sheet growth during the Last Glacial Maximum. *Nature* **370**, 357–359 (1994).
- Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–148 (1995).
- Tziperman, E. Inherently unstable climate behaviour due to weak thermohaline ocean circulation. *Nature* **386**, 592–595 (1997).
- Maier-Reimer, E. & Mikolajewicz, U. in *Oceanography* (eds Ayala-Castanares, A. *et al.*) 87–99 (UNAM Press, Mexico, 1989).
- Schmitz, W. J. Jr On the interbasin-scale thermohaline circulation. *Rev. Geophys.* **33**, 151–173 (1995).
- Petit, J. R., Briat, M. & Royer, A. Ice age aerosol content from East Antarctic ice core samples and past wind strength. *Nature* **293**, 391–394 (1981).
- Fairbanks, R. G. A 17,000-year glacial-eustatic sea level record: Influence of glacial meltwater rates on Younger Dryas event and deep ocean circulation. *Nature* **342**, 637–642 (1989).
- McIntyre, A. *et al.* in *Investigations of Late Quaternary Paleoclimatology and Paleoclimatology* (eds Cline, R. M. & Hays, J. D.) 43–76 (Geol. Soc. Am., Boulder, CO, 1976).
- Boyle, E. & Keigwin, L. D. North Atlantic thermohaline circulation during the past 20,000 years linked to high-latitude surface temperature. *Nature* **330**, 35–40 (1987).
- Kattenberg, A. *et al.* in *Climate Change 1995: the Science of Climate Change* (eds Houghton, J. T. *et al.*) 285–357 (Cambridge Univ. Press, 1996).

Acknowledgements. All numerical computations were conducted locally on a suite of IBM RS6000s including two SP2s. We thank S. Rahmstorf and T. Stocker for comments on an earlier version of this manuscript, and A. Broccoli for a review which significantly improved the manuscript. This work was supported by NSERC, CSHD, CICS, NOAA and Steacie operating grants, as well as an IBM SUR grant.

Correspondence and requests for materials should be addressed to A.J.W. (e-mail: weaver@ocean.seos.uvic.ca).

Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons

Tatyana V. Pestova*†, Sergei I. Borukhov* & Christopher U. T. Hellen*

* Department of Microbiology and Immunology, State University of New York Health Science Center at Brooklyn, 450 Clarkson Avenue, Brooklyn, New York 11203, USA

† A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119899 Moscow, Russia

The scanning model of translation initiation is a coherent description of how eukaryotic ribosomes reach the initiation codon after being recruited to the capped 5' end of messenger RNA. Five eukaryotic initiation factors (eIF 2, 3, 4A, 4B and 4F) with established functions have been assumed to be sufficient to mediate this process. Here we report that eIF1 and eIF1A are also both essential for translation initiation. In their absence, 43S ribosomal preinitiation complexes incubated with ATP, eIF4A, eIF4B and eIF4F bind exclusively to the cap-proximal region but are unable to reach the initiation codon. Individually, eIF1A enhances formation of this cap-proximal complex, and eIF1 weakly promotes formation of a 48S ribosomal complex at the initiation codon. These proteins act synergistically to mediate assembly of ribosomal initiation complexes at the initiation codon and dissociate aberrant complexes from the mRNA.

The ribosomal scanning model describes the basic steps of translation initiation on most eukaryotic mRNAs^{1,2}. In this process, a 43S complex, consisting of a ribosomal 40S subunit, eIF3 and an eIF2–GTP–initiator tRNA complex, binds mRNA at its 5' end and scans downstream until it locates the initiation codon. First, eIF4F binds the capped 5' end of the mRNA and, with eIF4A and eIF4B, creates an unstructured cap-proximal binding site for the 43S complex. This complex scans to the first downstream AUG triplet, which acts as the initiation codon. eIF5 stimulates GTP hydrolysis and release of factors from the resulting 48S complex, leaving the initiator tRNA in the P-site of the 40S subunit. The ribosomal 60S subunit then joins the 40S subunit and protein synthesis begins. Other factors, including eIF1 and eIF1A, have been implicated in the initiation of translation but their function remains obscure³.

Ribosomal binding to the end of an mRNA does not position it at the initiation codon, which is usually 50–100 nucleotides away. The 43S complex is thought to scan downstream, searching for the initiation codon. This model poses three basic questions. (1) Which factors are required for attachment of 43S complexes to capped mRNAs? (2) How does the 43S complex move on the mRNA, and which factors are required for this process? (3) How do components of the 43S complex interact with and inspect mRNA during scanning to recognize and reject mismatched interactions between triplets in the mRNA and the anticodon of initiator tRNA before the correct initiation codon is selected?

We have developed methods to reconstitute initiation from purified components and accurately to map the resulting initiation complexes on mRNAs^{4–6}. Here we have reconstituted early stages in initiation on natural capped β -globin mRNA and identified essential, unanticipated activities of eIF1 and eIF1A. eIF1, eIF1A, eIF4A, eIF4B and eIF4F are sufficient for 43S complexes to bind capped mRNAs and to form 48S complexes at the initiation codon. When eIF1 and eIF1A were omitted, 43S complexes bound near the 5' cap but did not reach the initiation codon. eIF1A enhanced the formation of these 5'-terminal complexes in the presence of the other five factors; in their presence, eIF1 slightly stimulated 48S complex formation and dissociation of aberrant 5' terminal complexes. These factors have distinct, synergistic activities that are required together for 48S complex assembly at the initiation codon.

Ribosome recruitment of capped mRNA

Ribosomal 48S complexes were assembled *in vitro* on β -globin

mRNA using purified factors (Fig. 1a). The position of 40S subunits on the mRNA in these complexes as mapped by toeprinting, which involves extension by reverse transcriptase of a primer annealed to a template RNA to which a ribosome is also bound. Synthesis of complementary DNA is arrested by the bound complex, yielding a toeprint at its leading edge that can be located on a sequencing gel. 48S complexes assembled on β -globin mRNA yield stops 15, 16 and 17 nucleotides downstream of the initiation codon⁷.

A ribosomal 'complex I', assembled from 40S subunits, initiator tRNA, eIF2, eIF3, eIF4A, eIF4B and eIF4F, yielded prominent toeprints 21–24 nucleotides from the 5' end of the mRNA (Fig. 2a, lane 3). Complex I did not form if 40S subunits, initiator tRNA, eIF2, eIF3 or eIF4F individually, or eIF4A, eIF4B and eIF4F together, were omitted, or if ATP was substituted by AMP–PNP (Fig. 2a, lanes 1–3, 2b, lanes 1–7, 10). The formation of complex I was greatly increased by eIF4B (Fig. 2b, lane 8). 43S complexes and eIF4A, eIF4B and eIF4F are therefore unable to form 48S complexes, and instead form ribosomal complexes near the 5' terminal cap. Parallel experiments using α -globin mRNA led to an identical conclusion. Ribosomal complexes yielded toeprints 16 and 23 nucleotides from the 5' end of this mRNA, but not at the initiation codon (data not shown).

Assembly of 48S complexes

48S complexes assemble correctly in rabbit reticulocyte lysate (RRL)⁷, which we therefore used as a source from which to purify additional factor(s) required for assembly of the 48S complex. The 0.5 M KCl ribosomal salt wash was divided into 0–40%, 40–50% and 50–70% ammonium sulphate precipitation fractions. The 50–70% fraction contained most of the activity that promoted assembly of a 48S complex (complex II) at the initiation codon on addition to reactions that contained 43S complexes and eIF4A, eIF4B and eIF4F (Fig. 2a, lane 4). This fraction was separated by elution from DEAE cellulose into 0.1 M KCl and 0.25 M KCl fractions that together had the same activity as the starting material (Fig. 1b). The 0.25 M KCl fraction doubled complex I formation but did not promote complex II formation; inclusion of the 0.1 M KCl fraction in similar reactions yielded small amounts of complex II without significantly altering formation of complex I, the main product (data not shown). The active constituents in these fractions were purified by chromatography (Fig. 1b) and were assayed after mixing, using toeprinting after each step to identify fractions and proteins that

promoted complex II formation. Apparently homogenous proteins of relative molecular mass (M_r) 13,500 and 19,000 (13.5K and 19K) (Fig. 1c, lanes 6, 7) were necessary, and together (but not individually) were sufficient for 43S complexes and eIF4A, eIF4B and eIF4F to form complex II without any trace of complex I (Fig. 2a, lanes 5–7). The amino-terminal sequence of the 19K protein was PKNKGKG, identical to that of rabbit eIF1A (ref. 8). The sequences of two tryptic peptides derived from the 13.5K protein were GDDLLPAGT and TLTTVQGIA, which correspond exactly to amino acids 18–26 and 45–52 of human eIF1 (refs 9, 10).

Including eIF1A in reactions lacking eIF1 increased the formation of complex I without forming complex II (Fig. 2a, lane 6). Conversely, eIF1 in the absence of eIF1A slightly reduced the prominence of complex I and yielded some complex II (Fig. 2a, lane 5). Without eIF1A, complex II yielded two toeprints 16–17 nucleotides downstream of the initiation codon, whereas complex II assembled with eIF1 and eIF1A yielded a third toeprint at 15 nucleotides downstream (Fig. 2a, lanes 5, 7). eIF1 and eIF1A were also required for complex II assembly on α -globin mRNA (data not shown). When complex I was assembled on β -globin mRNA without eIF1 and eIF1A, and they were added 5 min later, complex I disappeared completely but complex II formed as if eIF1 and eIF1A had been present throughout (Fig. 2a, lane 8). Complex I is therefore not a stable dead-end.

In toeprinting assays, the magnesium concentration was increased to 8 mM after the assembly reaction to optimize reverse

transcription. These ionic conditions can stabilize non-enzymatic binding of oligoribonucleotides to 40S subunits. 40S subunits and all factors required for 48S complex formation were incubated in a reaction mixture that contained 8 mM magnesium acetate from the beginning. Neither complex I nor II formed under these conditions (Fig. 2a, lane 9); they are therefore not simply the consequence of the magnesium-induced association of mRNA and the 40S subunit.

Recombinant eIF1 and eIF1A (Fig. 1c) were used to confirm that these factors are sufficient for 48S complex assembly when used with 43S complexes and eIF4A, eIF4B and eIF4F. Inclusion of recombinant eIF1 with these components yielded complex I and small amounts of complex II; recombinant eIF1A in place of eIF1 slightly increased complex I formation without formation of complex II (Fig. 3a, lanes 3, 4). Inclusion of recombinant eIF1 and eIF1A in similar reactions yielded complex II but not complex I (Fig. 3a, lane 6). Neither complex I nor complex II formed if 40S subunits, eIF2 or eIF3 individually, or eIF4A, eIF4B and eIF4F together, were omitted (Fig. 3a, lanes 7–10). The factors required for complex II formation and its position on mRNA indicate that it is a bona fide 48S complex. eIF1 and eIF1A are not required for recruitment of

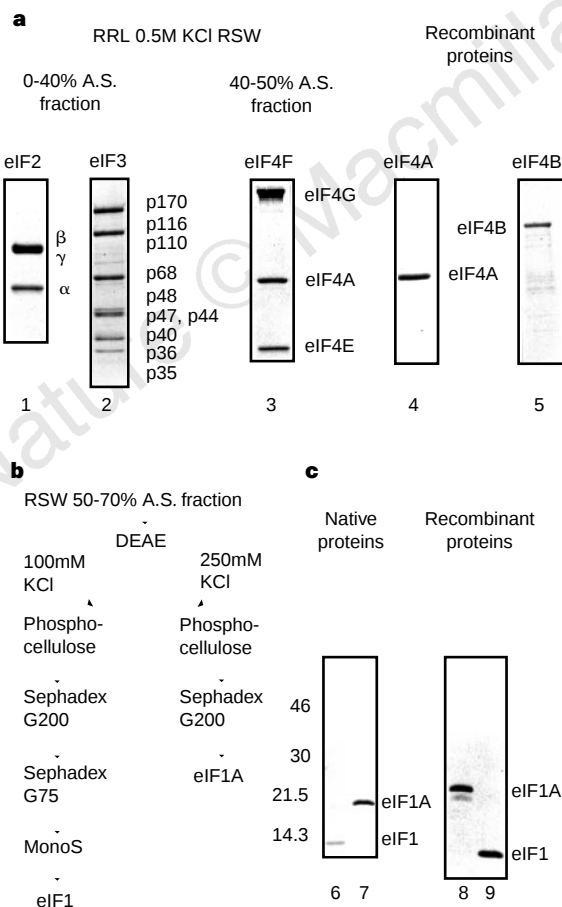


Figure 1 Composition and purification of proteins used in translation initiation. **a, c**, Overview of proteins used in reconstituting translation initiation. **b**, Purification scheme for eIF1 and eIF1A. SDS-PAGE gels were stained with Coomassie blue. The positions of molecular weight markers are indicated to the left of lane 6. Subunits of eIF2, eIF3 and eIF4F are indicated to the right of lanes 1–3.

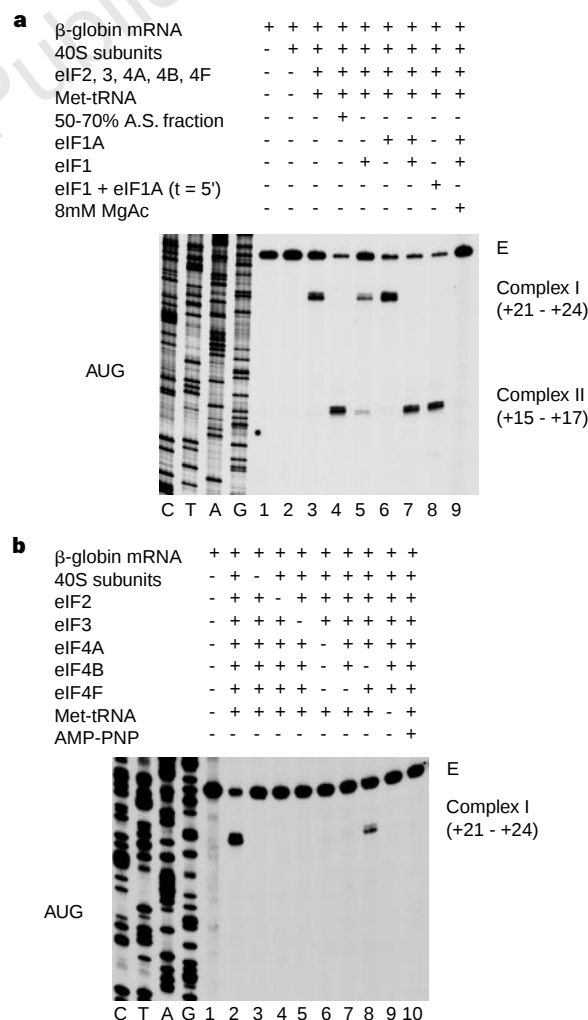
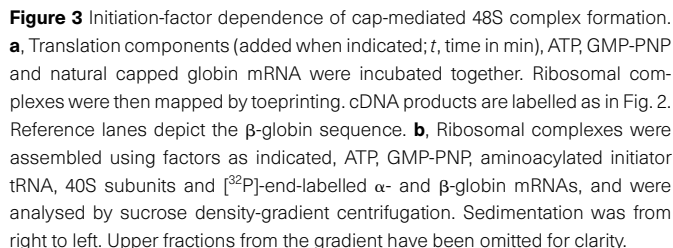


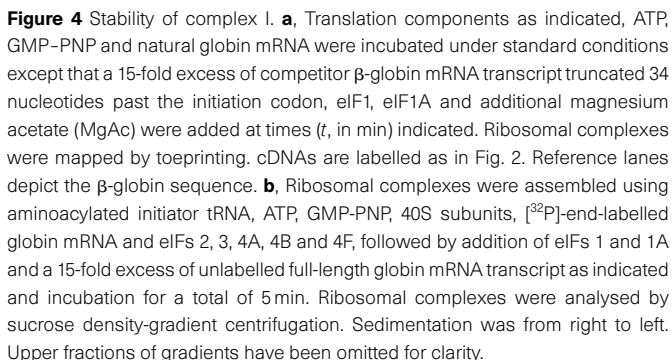
Figure 2 Assembly and toeprint analysis of ribosomal complexes on β -globin mRNA. **a, b**, Reaction mixtures contained ATP and GMP-PNP in addition to translation components, added when indicated (t, time in min). Full-length cDNA is marked E. cDNA products labelled 'Complex I (+21–+24)' and 'Complex II (+15–+17)' terminated 21–24 nucleotides from the 5' end and 15–17 nucleotides downstream of the initiation codon of β -globin mRNA, respectively. The position of the initiation codon is shown to the left of the reference lanes, which show the β -globin sequence derived using the same primer.

The influence of eIF1 and eIF1A on binding capped, 32 P-labelled globin mRNA to 40S subunits was assessed by sucrose density-gradient centrifugation. The incorporation of mRNA into ribosomal complexes was not influenced by eIF1A alone, but was greatly increased by eIF1 and even more by both factors in reactions that contained 43S complexes and eIF4A, eIF4B and eIF4F (Fig. 3b). Toeprinting indicated that the amounts of complex I formed without eIF1 and eIF1A, and of complex II formed in their presence, are similar, whereas much less complex I than complex II was detected by sucrose density-gradient centrifugation. Complex I is therefore unstable under conditions of sucrose density-gradient centrifugation.

The stability of complex I was assessed by toeprinting after the addition of competitor β -globin mRNAs that lacked the primer binding site. Inclusion of a 15-fold excess of competitor in reactions at the beginning of incubation abolished detectable formation of complex I; complex I was barely detectable when the same competitor excess was added after 5 min incubation (Fig. 4a, lanes 7, 8). In



Neither complex I nor complex II formed in reactions that contained eIF1, eIF1A and 8 mM Mg^{2+} (Figs 2a, lane 9 and 3a, lane 12). Unlike complex II, complex I can assemble at normal (2 mM) and elevated (8 mM) Mg^{2+} concentrations in the absence of eIF1 and eIF1A (Fig. 4a, lanes 2, 3). The failure to form complex I in the presence of 8 mM Mg^{2+} , eIF1 and eIF1A could be because these factors prevent its formation under these conditions, or because they destabilize it immediately. We investigated the effect of these factors on preformed complex I. Neither factor individually affected its prominence if added to reactions when the Mg^{2+} concentration was increased from 2 to 8 mM but, if these factors were added together, complex I was not detectable (Fig. 4a, lanes 4–6). eIF1 and eIF1A can therefore identify and promote the dissociation of incorrectly assembled complex I under conditions that do not allow the formation of complex II (that is, 8 mM Mg^{2+}). These



factors might detect that complex I is assembled incorrectly and destabilize it, or complex I might be intrinsically unstable, cycling on and off the mRNA, and could be trapped in the 'off' state by these factors.

The fate of complex I

Complex I can be dissociated from mRNA by eIF1 and eIF1A at Mg^{2+} concentrations that do not allow complex II to form. At normal Mg^{2+} concentrations, addition of these factors to preformed complex I might result in it either migrating to the initiation codon under their influence, or dissociating from the mRNA spontaneously or under their influence, so the 43S complex can restart the initiation process.

To distinguish between these possibilities, sucrose density-gradient centrifugation was used to resolve 48S complexes assembled on labelled globin mRNA with and without unlabelled competitor. If eIF1 and eIF1A promote migration of complex I to the initiation codon on the same mRNA, simultaneous addition of excess unlabelled mRNA and these factors to complex I should yield 48S complexes on the labelled mRNA. Alternatively, if eIF1 and eIF1A promote the formation of 48S complex on another mRNA in a new cycle of initiation then competition should yield 48S complexes on the excess unlabelled mRNA that would not be detected by scintillation counting of sucrose density-gradient fractions. A prominent 48S peak was detected on labelled mRNA in the absence of competitor if eIF1 and eIF1A were present from the beginning (Fig. 4b). Simultaneous addition of 43S complexes, eIF1, eIF1A, eIF4A, eIF4B and eIF4F and a 15-fold excess of unlabelled competitor over labelled globin mRNA to reactions prevented 48S complex assembly on the labelled mRNA. Addition of eIF1, eIF1A and excess competitor to similar reactions 5 min after the start of complex I assembly did not appreciably increase the assembly of 48S complex on the labelled mRNA. This finding suggests that complex I is not an immediate precursor of 48S complexes, and that it dissociates either spontaneously or under the influence of eIF1 and eIF1A, releasing 43S complexes that enter another round of initiation. However, we cannot rule out the possibility that the rate of spontaneous dissociation of complex I is greater than the rate of its fixation by eIF1 and eIF1A in a form that leads to 48S complex assembly. In this circumstance, preassembly of complex I before addition of eIF1, eIF1A and competitor mRNA would not favour 48S complex formation on the same mRNA.

Destabilization of aberrant ribosomal complexes

Initiation on some eukaryotic mRNAs is cap independent, resulting instead from internal ribosomal entry. Encephalomyocarditis virus (EMCV) and classical swine fever virus (CSFV) exemplify different internal initiation mechanisms; neither involves scanning or requires eIF1 or eIF1A for 48S complex assembly⁴⁻⁶.

EMCV translation initiates at AUG₈₃₄ and infrequently at AUG₈₂₆ (ref. 11). 43S complexes and eIF4A, eIF4B and eIF4F form 48S complexes at AUG₈₂₆ and predominantly at AUG₈₃₄ (refs 4, 5) (Fig. 4a, lane 2). The proportion of complexes assembled at AUG₈₂₆ exceeds the relative frequency of initiation at this codon during translation. eIF1 strongly reduced 48S complex formation at AUG₈₂₆ and increased 48S complex formation at AUG₈₃₄ commensurately; eIF1A enhanced the stop 15 nucleotides downstream of AUG₈₃₄ but had no effect on the relative proportion of complexes formed at AUG₈₂₆ and AUG₈₃₄. Together these factors had an additive effect (Fig. 5a, lanes 3–5). eIF1 and eIF1A were present throughout these reactions so these results do not discriminate between the possibilities that these factors prevent the assembly of 48S complex at AUG₈₂₆ and that they destabilize complexes that have already assembled at this triplet. Addition of both factors to reactions after 5 min incubation reduced 48S complex formation at AUG₈₂₆ as much as when they were present from the beginning (Fig. 5a, lanes 5, 6). We do not know which features of the complex

assembled at AUG₈₂₆ target it for destabilization. eIF1 and eIF1A can therefore destabilize incorrectly assembled ribosomal complexes and increase 48S complex assembly at the initiation codon during IRES- and cap-mediated modes of initiation. 48S complexes assembled equally well at EMCV AUG₈₃₄ at 2 mM and at 8 mM Mg^{2+} , with or without eIF1 and eIF1A, in otherwise identical conditions (Fig. 5a). This difference between IRES- and cap-mediated initiation suggests that scanning may be inhibited at high magnesium concentrations.

48S complex formation on the CSFV IRES requires only eIF2 and eIF3 (ref. 6). This IRES promotes direct binding of 43S complexes to the initiation codon, yielding toeprints at UGA₃₉₀₋₃₉₂. eIF1 and eIF1A did not affect the yield of 48S complexes but slightly altered the toeprint pattern at UGA₃₉₀₋₃₉₂ (Fig. 5b, lanes 2, 3). 48S complexes assembled equally well on the CSFV IRES at 2 and 8 mM Mg^{2+} (data not shown), as described for EMCV.

Discussion

We have found that eIF1 and eIF1A are essential for cap-mediated initiation of translation in higher eukaryotes. They have distinct, synergistic activities that together promote 48S complex assembly at the initiation codon and dissociation of aberrant ribosome-mRNA complexes. Without them, 43S complexes bound exclusively to the cap-proximal 5' non-translated region in a reaction that required ATP, eIF4A, eIF4B and eIF4F. These ribosomal complexes did not reach the initiation codon. When eIF1 and eIF1A were added, the bound ribosomal complex dissociated from the mRNA and entered

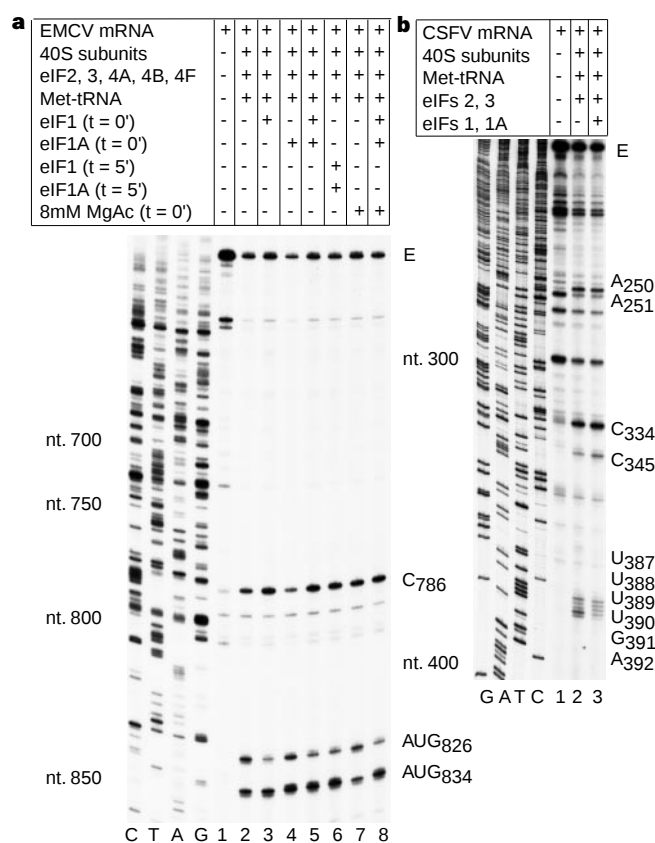


Figure 5 Assembly and toeprint analysis of ribosomal complexes on EMCV and CSFV mRNAs. **a**, EMCV; **b**, CSFV. Reaction mixtures contained ATP and GMP-PNP, and translation components were added when indicated (time *t*, in min). Full-length cDNAs are marked E. cDNA products labelled AUG₈₃₄ and AUG₈₂₆ terminated 15–17 nucleotides (nt.) from the stated EMCV codon. The cDNA product labelled C₇₈₆ terminated at this nucleotide⁴. cDNA products labelled A₂₅₀, C₃₃₄, G₃₄₅ and C₃₈₇–A₃₉₂ terminated at these nucleotides⁵. The positions of some EMCV and CSFV nucleotides are indicated on the left.

a new round of initiation that led to 48S complex formation. The roles in initiation of eIF1 and eIF1A have remained obscure since their identification^{12,13}. They have been described as having pleiotropic and usually minor effects on various stages in initiation^{14–18}, but homologous factors are essential for the viability of yeast *Saccharomyces cerevisiae*^{19,20}. By using toeprinting to map ribosomal complexes on mRNAs (rather than sucrose density-gradient centrifugation merely to monitor their association), we have now identified essential, unanticipated activities of eIF1 and eIF1A.

Ribosomal scanning is the most coherent model for cap-mediated initiation of translation, and we shall discuss our data in terms of it. Complex I stopped primer extension at identical 5'-proximal positions on mRNA regardless whether or not it was assembled under ionic conditions that allowed scanning. It probably did not move on the mRNA after attachment. eIFs 2, 3, 4A, 4B and 4F are therefore sufficient for attachment of a ribosome to capped mRNA but not for it to begin scanning. We do not know how the 43S complex interacts with mRNA during scanning, but conclude that these five factors do not provide the 43S complex with the necessary 'active centre'. Inhibition of scanning by raised concentrations of Mg²⁺ may be due to alterations in the conformation of this active centre. Little complex II was formed in the presence of eIF1 unless eIF1A was also present. eIF1A thus increases the competence of 43S complexes to scan, possibly by increasing the affinity of eIF1 for the 43S-mRNA complex or by enhancing its stability on an mRNA during scanning. eIF1A is a strong RNA-binding protein²¹. The mRNA-binding site on eukaryotic 40S subunits appears to be an open cleft that can bind internal sequences of an mRNA in the absence of a free 5' end during IRES-mediated initiation²². This cleft probably also binds an internal sequence downstream of the eIF4F-bound 5'-cap during cap-mediated initiation, but may be clamped shut during scanning so the mRNA can be inspected base by base without bypassing hairpins that might harbour the initiation codon²³. eIF1 and eIF1A are candidates to close this cleft.

eIF1A increased the formation of complex I in reactions lacking eIF1. We have not distinguished between the possibilities that eIF1A stabilizes 43S complexes in solution and that it stabilizes their interaction with mRNA^{17–20}. Gel filtration analysis showed that eIF1 stabilizes binding of eIF1A to 43S complexes¹⁶. eIF1A strongly enhanced eIF1-mediated assembly of 48S complexes and caused a striking qualitative difference in the resulting complex from that produced in its absence. 48S complexes formed in the presence of eIF1 and eIF1A yielded prominent toeprints 15–17 nucleotides, downstream of the initiation codon, whereas in the presence of eIF1 alone they yielded only the stops 16–17 nucleotides downstream. Similarly, a third toeprint 15 nucleotides downstream of EMCV AUG₈₃₄ was apparent only when eIF1A was present. We have not yet determined whether this difference is indicative of a conformational change in the 48S complex induced by eIF1A that is necessary for subunit joining to form an 80S complex.

A second critical function of eIF1 and eIF1A in initiation is to maintain the accuracy of this process by recognizing and destabilizing aberrant preinitiation complexes. eIF1 alone can recognize and destabilize ribosomal complexes incorrectly assembled at EMCV AUG₈₂₆. This editing function is also active on capped mRNAs but depends on the presence of eIF1A. Together, these factors destabilize complex I. We do not know which properties of this complex and of the complex at EMCV AUG₈₂₆ target them for destabilization by eIF1. However, identification of this activity in mammalian eIF1 is consistent with properties of its yeast homologue Sui1, which is a monitor of translational accuracy. Mutations in Sui1 allow aberrant initiation at a UUG triplet^{19,24}. The specific role of Sui1 in maintaining translation accuracy is not known. Our finding that eIF1 destabilizes aberrant ribosomal complexes suggest that these Sui1 mutants may be unable to discriminate against mismatched codon-anticodon interactions. The importance of these activities of eIF1

and eIF1A is to monitor the accuracy of the start-site selection throughout the scanning process to prevent spurious initiation events.

Taken together, our results suggest a model for cap-mediated initiation in which interactions between eIF3, mRNA and the eIF4G subunit of cap-bound eIF4F promote attachment of 43S complexes to unstructured, cap-proximal regions of an mRNA. eIF1 and eIF1A must interact with 43S complexes when or immediately after they bind mRNA for them to be competent to begin scanning. These factors may contribute to the correct interactions of the 43S complex with mRNA that enable it to enter a processive mode, for example by forming part of the channel through which mRNA moves during ribosomal scanning, by contributing to the ribosomal translocation process, or by both of these. Indeed, these activities could be integrated as part of a mechanism for base-by-base inspection of an mRNA during scanning. □

Methods

Plasmids. pCSFV (1–442).NS' (ref. 6), pTE1 (ref. 25), pBS⁺(β-globin) (ref. 26), pT7-7-1A (ref. 21), pET (His₆-eIF4A) and pET (His₆-eIF4B) (refs 4, 5) have been described. cDNA corresponding to the eIF1 coding region⁹ was amplified by the polymerase chain reaction (PCR) from phagemid ATCC 156441 from the American Type Culture Collection (Rockville, MD) with primers 5'-CGACGGATCCTATGTCCGCTATCCAG-3' and 5'-CTAGAAGCTTAAACCATGAACC-3' and inserted into pQE31 (Qiagen Santa Clarita, CA) to yield pQE(His₆-eIF1). cDNA corresponding to the eIF1A coding region was amplified by PCR from pT7-7-1A (ref. 21) with primers 5'-ATATGGATCGATGCCCAAGAATAAAGGAGG-3' and 5'-CAACAAGCTTGATGATGT-CATCAATATCTTC-3' and inserted into pET28 (Novagen, Madison, WI) to yield pET (His₆-eIF1A).

Purification of factors and 40S ribosomal subunits. 40S ribosomal subunits and eIF2, eIF3 and eIF4F were purified from RRL (Green Hectares, Oregon, WI) as described^{4,6}. No cross-contamination was detected by western blotting using specific antibodies and functional assays dependent on each factor^{4–6}. eIF1 and eIF1A were purified from the 50–70% ammonium sulphate ribosomal salt wash precipitation fraction from 2 litres of RRL. This fraction was resuspended in buffer A (20 mM Tris-HCl, pH 7.5, 10% glycerol, 1 mM DTT, 0.1 mM EDTA) containing 100 mM KCl, applied to a DEAE cellulose (Whatman DE52) column and eluted first with buffer A containing 100 mM KCl and then with buffer A containing 250 mM KCl. These fractions were the starting materials for purification of eIF1 and eIF1A, respectively. eIF1 in the 400–600 mM KCl fraction eluted from a phosphocellulose (Whatman P11) column was applied in buffer A containing 200 mM KCl to FPLC Superdex G200 and then G75 gel filtration columns (Pharmacia). The eIF1-containing fraction in buffer B (20 mM HEPES, pH 7.5, 10% glycerol, 1 mM DTT, 0.1 mM EDTA) containing 100 mM KCl was applied to an FPLC MonoS HR 5/5 column (Pharmacia). Fractions were collected across a 100–500 mM KCl gradient; apparently homogeneous eIF1 eluted with 250 mM KCl. eIF1A in the 700–900 mM KCl elution fraction from a phosphocellulose column was purified to apparent homogeneity by gel filtration in buffer A containing 200 mM KCl on an FPLC Superdex G200 column. Recombinant eIF4A and eIF4B were purified as described⁴. Recombinant eIF1 and eIF1A were expressed in *Escherichia coli* BL21 (DE3) and purified using Ni²⁺-NTA (Qiagen) and heparin-Sepharose (Pharmacia) matrices.

Sequencing of eIFs 1 and 1A. Purified 13.5K and 19K proteins were resolved by gel electrophoresis. The 19K protein was transferred to PVDF membrane for N-terminal sequencing done using an Applied Biosystems Procise sequencer. Gel bands containing the 13.5K protein were excised, washed with 0.2 M ammonium bicarbonate/50% acetonitrile, dried and rehydrated in 0.2 M ammonium bicarbonate. Overnight digestion by sequencing grade modified trypsin (Promega, Madison, WI) was stopped by addition of 10% trifluoroacetic acid (TFA) to a final concentration of 1%. The supernatant obtained after washing the gel piece with 0.1% TFA/60% acetonitrile was reduced by rotary evaporation. Peptides were isolated from this extract by reversed-phase liquid chromatography, and two well-resolved peaks were sequenced.

Assembly and analysis of ribosomal complexes. For analysis of 48S complexes by sucrose density-gradient centrifugation, 0.3 μg native α- and β-

globin mRNAs (Life Technologies, Grand Island, NY) that had been 3'-end-labelled with [³²P]pCp using T4 RNA ligase was incubated for 5 min at 30 °C in a 100-μl reaction volume that contained buffer C (2 mM DTT, 100 mM potassium acetate, 20 mM Tris, pH 7.6) with 2 mM magnesium acetate, 1 mM ATP, 0.1 mM guanylimidodiphosphate (GMP-PNP), 0.25 mM spermidine, eIF2 (6 μg), eIF3 (14 μg), eIF4A (4 μg), eIF4B (1 μg), eIF4F (4 μg), 12 pmol Met-tRNA^{Met} and 12 pmol 40S subunits, with or without eIF1 (1 μg), eIF1A (1 μg) and unlabelled competitor mRNA, as indicated in the text. 48S and ribonucleoprotein complexes were resolved by centrifugation through sucrose density gradients as described^{4,6}. For toeprint analysis, ribosomal complexes were assembled essentially as described above by incubating 1 μg CSFV or EMCV mRNA⁴⁻⁶ or 0.3 μg native α- and β-globin mRNAs for 5 min at 30 °C in a 40-μl reaction volume. In some experiments, reaction mixtures were supplemented with 4–8 μg of different RRL subfractions as indicated in the text. Incubation was continued for 3 min at 30 °C following addition of 4 pmol primer 5'-GCATTTCAGAGGACAGG-3' (complementary to β-globin nucleotides 177–194), 5'-GTCAATAACTCTCTGG-3' (complementary to EMCV nucleotides 957–974) or 5'-CTCGTTTGGCGACATGCC-3' (complementary to part of the NS' coding sequence in CSFV-NS' mRNA) as appropriate. Ribosomal and RNP complexes were analysed by primer extension^{4,5} using avian myeloblastosis virus reverse transcriptase (Promega) in the presence of [α-³²P]dATP (~6,000 Ci mmol⁻¹; ICN Radiochemicals). cDNA products were analysed by electrophoresis through 6% polyacrylamide sequencing gel. cDNA products were compared with appropriate dideoxynucleotide sequence ladders.

Received 11 March; accepted 23 June 1998.

1. Kozak, M. How do eukaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123 (1978).
2. Kozak, M. The scanning model for translation: An update. *J. Cell. Biol.* **108**, 229–241 (1989).
3. Merrick, W. C. & Hershey, J. W. B. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B. & Sonenberg, N.) 31–70 (Cold Spring Harbor Laboratory Press, NY, 1996).
4. Pestova, T. V., Hellen, C. U. T. & Shatsky, I. N. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell. Biol.* **16**, 6859–6869 (1996).
5. Pestova, T. V., Shatsky, I. N. & Hellen, C. U. T. Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol. Cell. Biol.* **16**, 6870–6878 (1996).
6. Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. T. A prokaryotic-like mode of binding of cytoplasmic eukaryotic ribosomes to the initiation codon during internal initiation of translation of Hepatitis C and Classical Swine fever virus RNAs. *Genes Dev.* **12**, 67–83 (1998).
7. Anthony, D. D. & Merrick, W. C. Analysis of 40S and 80S complexes with mRNA as measured by sucrose density gradients and primer extension inhibition. *J. Biol. Chem.* **267**, 1554–1562 (1992).

8. Dever, T. E. *et al.* Determination of the amino acid sequence of rabbit, human, and wheat germ protein synthesis factor eIF-4C by cloning and chemical sequencing. *J. Biol. Chem.* **269**, 3212–3218 (1994).
9. Fields, C. & Adams, M. D. Expressed sequence tags identify a human isolog of the *sui1* translation initiation factor. *Biochem. Biophys. Res. Commun.* **198**, 288–291 (1994).
10. Kasperaitis, M. A. M., Voorma, H. O. & Thomas, A. A. M. The partial amino acid sequence of eukaryotic translation initiation factor 1 and its similarity to yeast initiation factor *sui1*. *FEBS Lett.* **365**, 47–50 (1995).
11. Kaminski, A., Howell, M. T. & Jackson, R. J. Initiation of encephalomyocarditis virus RNA translation: the authentic initiation site is not selected by a scanning mechanism. *EMBO J.* **9**, 3753–3759 (1990).
12. Kemper, W. M., Berry, K. W. & Merrick, W. C. Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2Bα and M2Bβ. *J. Biol. Chem.* **251**, 5551–5557 (1976).
13. Schreier, M. H., Erni, B. & Staehelin, T. Initiation of mammalian protein synthesis. I. Purification and characterization of seven initiation factors. *J. Mol. Biol.* **116**, 727–753 (1977).
14. Trachsel, H., Erni, B., Schreier, M. H. & Staehelin, T. Initiation of mammalian protein synthesis. II. The assembly of the initiation complex with purified initiation factors. *J. Mol. Biol.* **116**, 755–767 (1977).
15. Benne, R. & Hershey, J. W. B. The mechanism of action of protein synthesis initiation factors from rabbit reticulocytes. *J. Biol. Chem.* **253**, 3078–3087 (1978).
16. Thomas, A., Goumans, H., Voorma, H. O. & Benne, R. The mechanism of action of eukaryotic initiation factor 4C in protein synthesis. *Eur. J. Biochem.* **107**, 39–45 (1980).
17. Thomas, A., Spaan, W., van Steeg, H., Voorma, H. O. & Benne, R. Mode of action of protein synthesis initiation factor eIF-1 from rabbit reticulocytes. *FEBS Lett.* **116**, 67–71 (1980).
18. Chaudhuri, J., Si, K. & Maitra, U. Function of eukaryotic translation initiation factor 1A (eIF1) (formerly called eIF-4C) in initiation of protein synthesis. *J. Biol. Chem.* **272**, 7883–7891 (1997).
19. Yoon, H. & Donahue, T. F. The *sui1* suppressor locus in *Saccharomyces cerevisiae* encodes a translation factor that functions during tRNA^{Met} recognition of the start codon. *Mol. Cell. Biol.* **12**, 248–260 (1992).
20. Wei, C.-L., MacMillan, S. E. & Hershey, J. W. B. Characterization of yeast translation initiation factor 1A and cloning of its essential gene. *J. Biol. Chem.* **270**, 5764–5771 (1995).
21. Wei, C.-L., Kainuma, M. & Hershey, J. W. B. Protein synthesis initiation factor eIF-1A is a moderately abundant RNA-binding protein. *J. Biol. Chem.* **270**, 22788–22794 (1995).
22. Chen, C.-Y. & Sarnow, P. Initiation of protein synthesis by the eukaryotic translation apparatus on circular RNAs. *Science* **268**, 415–417 (1995).
23. Jackson, R. J. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B. & Sonenberg, N.) 71–112 (Cold Spring Harbor Laboratory Press, NY, 1996).
24. Cui, Y., Dinman, J. D., Kinzy, T. G. & Peltz, S. W. The Mof2/Sui1 protein is a general monitor of translational accuracy. *Mol. Cell. Biol.* **18**, 1506–1516 (1998).
25. Evstafieva, A. G., Ugarova, T. Y., Chernov, B. K. & Shatsky, I. N. A complex RNA sequence determines the internal initiation of encephalomyocarditis virus RNA translation. *Nucleic Acids Res.* **19**, 665–671 (1991).
26. Hellen, C. U. T. *et al.* A cytoplasmic 57kDa protein that is required for translation of picornavirus RNA by internal ribosomal entry is identical to the nuclear pyrimidine tract-binding protein. *Proc. Natl Acad. Sci. USA* **90**, 7642–7646 (1993).

Acknowledgements. We thank W. Merrick and N. Sonenberg for discussion, L. Siconolfi-Baez for sequencing the eIF1 and eIF1A proteins, and H.-P. Vornlocher and J. W. B. Hershey for the p17-7-1A plasmid. These studies were supported by grants from the NIH (C.U.T.H. and S.I.B.) and the NSF (C.U.T.H.).

Correspondence and requests for materials should be addressed to T.V.P. (e-mail: tpestova@netmail.hscbklyn.edu).

KNOW YOUR COPY RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA. Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Old and blue white-dwarf stars as a detectable source of microlensing events

Brad M. S. Hansen

Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, Ontario M5S 3H8, Canada

The analysis¹ of gravitational microlensing events of stars^{2,3} in the Large Magellanic Cloud places the masses of the lensing objects in the range 0.3–0.8 solar masses, suggesting that they might be old white-dwarf stars. Such objects represent the last stage of stellar evolution: they are the cooling cores of stars that have lost their atmospheres after nuclear fusion has ceased in their centres. If white dwarfs exist in abundance in the halo of our Galaxy, this would have profound implications for our understanding of the early generations of stars in the Universe^{4–6}. Previous attempts to constrain theoretically^{6–8} the contribution of white dwarfs to microlensing indicate that they can account for only a small fraction of the events. But these estimates relied on models of white-dwarf cooling that are inadequate for describing the properties of the oldest such objects. Here I present cooling models appropriate for very old white dwarfs. I find, using these models, that the widely held notion that old white dwarfs are red applies only to those with a helium atmosphere; old white dwarfs with hydrogen atmospheres, which could be a considerable fraction of the total population, will appear rather blue, with colours similar to those of the faint blue sources in the Hubble Deep Field. Observational searches for the population of microlensing objects should therefore look for faint blue objects, rather than faint red ones.

The issue of the observability of old white dwarfs is inextricably linked to the age of the object, because white dwarfs fade with time. The primary uncertainty in the calculation of accurate cooling models is the description of energy transport in the stellar atmosphere, which affects both the cooling rate⁹ and observational appearance¹⁰. Old white dwarfs have luminosities L such that $\log(L/L_{\odot}) < -4$ and effective temperatures $T_{\text{eff}} < 6,000$ K (here $L_{\odot}$ is the solar luminosity). The atmospheric constituents at these temperatures are neutral, so that the primary opacity sources are the collisionally induced absorption of molecular hydrogen (for hydrogen atmospheres) or Rayleigh scattering (for helium atmospheres). The molecular opacities are strongly wavelength-dependent and

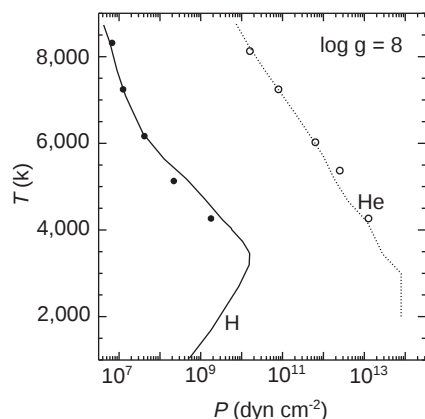


Figure 1 Location of the photosphere in white dwarfs. The solid and dotted lines indicate the evolution of the photospheres for hydrogen and helium atmospheres respectively. The solid and open circles are the same quantities from the work of Bergeron *et al.*¹⁰.

require a detailed radiative transfer calculation. The calculation of such atmosphere models¹⁰ has led to advances in the study of the basic physical parameters of old white dwarfs¹¹. However, previous cooling models have used outer boundary conditions calculated using simplified atmosphere calculations, so that the self-consistent determination of cooling ages from atmospheric models was not possible for the oldest white dwarfs. The cooling models described here aim to rectify that uncertainty.

My calculations use a white-dwarf cooling model originally developed for the study of the companions to millisecond pulsars¹². To this I have added a detailed atmospheric model using the Feautrier and Avrett-Krook methods¹³ to solve the radiative transfer at the surface. The atmospheric model provides an outer boundary condition (taken to be the photosphere) for the interior model which describes the cooling of the white dwarf. The position of the photosphere for helium and hydrogen atmospheres differs dramatically, because neutral helium does not form molecules so that the opacity κ is much lower and hence the density at the photosphere $\rho \propto 1/\kappa T_{\text{eff}}$ is much larger. Figure 1 shows the location of the photosphere for hydrogen and helium atmospheres, respectively. In both cases, the photospheric density increases as the star cools and T_{eff} drops. This trend ends when both atmospheres reach $T_{\text{eff}} \approx 3,000$ K, but for different reasons. For helium atmospheres, pressure ionization of helium increases the opacity dramatically, halting the inward motion of the photosphere. For hydrogen atmospheres, the molecular hydrogen opacity is stronger at long wavelengths $\lambda > 1 \mu\text{m}$, so that the opacity increases when the peak of the black-body $\lambda_{\text{ob}} \approx 1.7 \mu\text{m}/(T_{\text{eff}}/3,000 \text{ K})$ moves into that range, and hence the photospheric density moves outwards again. The molecular opacity is strongly wavelength-dependent, so that the observational appearance of these old stars deviates significantly from a black-body spectrum, in a similar fashion to that seen in brown-dwarf atmospheres¹⁴.

Figure 2 shows the comparison of the cooling results with the best models in the literature^{15,16}. The agreement with the hydrogen-atmosphere models is excellent until the models reach $T_{\text{eff}} \approx 6,000$ K, at which point my improved treatment of the

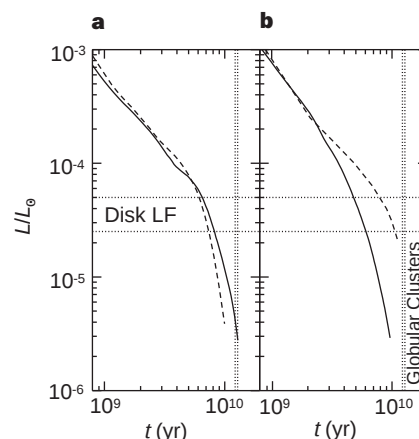


Figure 2 Calculated cooling curves. **a**, Comparison of this work (solid line) with the models of Wood¹⁵ (dashed line). The model is for a $0.6 M_{\odot}$ white dwarf with an oxygen core, and a hydrogen mass fraction of 10^{-4} . The improved boundary condition leads to a slight flattening of the cooling curve at $\log(L/L_{\odot}) \approx -4$ and longer cooling times thereafter. **b**, Comparison of this work (solid line) with the model of Salaris *et al.*¹⁶ (dashed line). The model is for a $0.6 M_{\odot}$ carbon/oxygen model with a helium mass fraction of $10^{-3.25}$. The much longer cooling of the Salaris model is because of the outdated atmospheric model used in that work. In both panels, the horizontal dotted lines enclose the location of the turnover in the disk white-dwarf luminosity function (LF). The vertical dotted lines indicate the mean age of the globular clusters²⁷, indicating the expected age of any halo white-dwarf population. The large variations in the different models at these ages indicate how important are accurate atmosphere models.

atmospheric physics results in slower cooling. The difference in the helium models is more dramatic, with the new models cooling rather more rapidly. This is a result of the extremely low opacity of neutral helium. Application of the new models to the white-dwarf luminosity function shows that previous age estimates for Galactic disk hydrogen dwarfs were reasonably accurate, although disk helium dwarf ages were somewhat overestimated. However, for white dwarfs residing in the Galactic halo or in globular clusters, the differences are very important. In particular, previous efforts to place constraints on white-dwarf dark matter^{6–8} have been unduly pessimistic because they use inappropriate white-dwarf models.

The observational constraints on local white-dwarf dark matter centre on two sources, the luminosity function determined from proper motions¹⁷ and searches for point sources in the Hubble Deep Field^{18–20}. Unfortunately, the proper-motion sample suffers from poorly constrained incompleteness at the fainter magnitudes (the luminosity function declines much more rapidly than the luminosity function based on white dwarfs in binaries²¹). Thus, inferring how many objects could have been seen at fainter magnitudes may lead to erroneous results. More robust estimates must rely on the Hubble Deep Field (HDF) alone. Figure 3 shows the point sources detected in the HDF by various workers, along with cooling tracks for both hydrogen and helium atmospheres. The first result that one can glean from this is that the HDF places no limits on helium-atmosphere dwarfs, because the rapid cooling means that they become unobservable within 6 Gyr, which is approximately half their expected age. The second and most interesting result is that the hydrogen-atmosphere dwarfs are potentially observable and would lie in the region of the faint blue objects that were indeed detected. This is contrary to the conventional wisdom which states that white dwarfs should become redder with age because that is the trend shown by a black body (and indeed that is what happens to the helium dwarfs). The deviation towards the blue shown by the hydrogen-atmosphere dwarfs is again a consequence of the strong molecular hydrogen opacity which also causes the photosphere to move to lower densities in Fig. 1. The strong opacity at longer wavelengths (red) forces the stellar flux to emerge at shorter

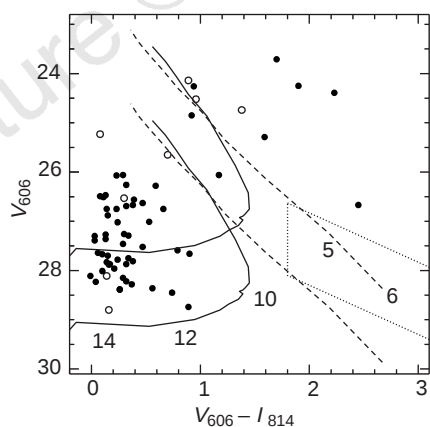


Figure 3 Stellar objects in the Hubble Deep Field. The filled circles are the unresolved objects detected by Elson *et al.*¹⁹; the open circles are the point sources from Mendez *et al.*²⁰. Although Flynn *et al.*¹⁸ did not publish a table of detections, the dotted line encloses their 'halo region', which they used to constrain the halo white-dwarf population. The dashed lines show the cooling behaviour of a helium atmosphere white dwarf at a distance of 1 kpc (upper line) and 2 kpc (lower line). The upper curve is labelled with ages in Gyr at appropriate points. The solid lines represent the corresponding evolution of a hydrogen-atmosphere dwarf at 1 and 2 kpc. The dwarf ages are shown on the lower curve in this case. The bandpasses are the Hubble Telescope bandpasses from Holtzman *et al.*²⁸. The blueward shift for ages 10–12 Gyr is due to the presence of molecular hydrogen in the atmosphere.

wavelengths (blue) where the opacity is smaller. Thus, hydrogen atmosphere white dwarfs may have already been detected in the halo.

To connect quantitatively these findings to the observational searches, one also requires a model for the relative populations of hydrogen- and helium-atmosphere white dwarfs. Recent advances in asteroseismological investigations of white dwarfs²² have confirmed that many hydrogen-atmosphere white dwarfs have hydrogen surface layers of mass $\sim 10^{-4} M_{\odot}$ (here $M_{\odot}$ is the solar mass). This suggests that such stars will survive as hydrogen-atmosphere dwarfs despite the mixing effects of surface convection zones. Although these estimates are for disk white dwarfs, the mechanisms for removing hydrogen from the surface of a white dwarf^{9,23,24} should become less efficient in stars with lower metallicity, such as those in the halo and in globular clusters. The relative proportion of hydrogen atmospheres amongst the cool white dwarfs in the disk appears to be $>50\%$, so I will adopt this as a conservative estimate. In the future, the effect of advanced age and lower progenitor metallicity could be empirically tested by examining the faint white dwarfs in globular clusters²⁵.

How many white dwarfs would one expect to see in the Hubble Deep Field? The HDF probes great distances, but only over a very small portion of sky, so that the total Galactic volume sampled is quite small. Given a limiting V-band magnitude $m_V \approx 28$ and an absolute magnitude for old white dwarfs $M_V \approx 17$, the corresponding volume probed is only:

$$V_{\text{HDF}} \approx (5.4 \text{ pc})^3 10^{0.6(m_V - 28) - (M_V - 17)}$$

Thus, even if the white dwarfs constitute the entire local dark matter ($\sim 0.01 M_{\odot} \text{ pc}^{-3}$; ref. 26), their space density is only $\sim 0.02 \text{ pc}^{-3}$ and hence we expect to find only three, even if all the white dwarfs have hydrogen atmospheres and are 12 Gyr old (an age appropriate for globular cluster and halo stars). Our most realistic estimate invokes only 50% of the total number in hydrogen-atmosphere white dwarfs and requires that only 50% of the dark matter is in the form of white dwarfs¹. In this case, the volume probed by the HDF is too small to reliably detect even one object. Nevertheless, as one can see in Fig. 3, several point sources have been detected in the HDF. Given the relative profusion of these objects, most must have another origin, such as extragalactic star-forming regions¹⁹.

These results have two implications for the search for dark matter. The first is that deep surveys for very red objects are unlikely to be successful in the near future because the red population (helium-atmosphere dwarfs) will be very faint. However, searches for faint blue objects could be much more successful. The second implication is that the hydrogen-atmosphere dwarfs are observable, but that the HDF results suggest that there are other unresolved populations of objects with similar colours and magnitudes. Thus determinations of proper motions will be essential, to distinguish Galactic from extragalactic objects. □

Received 30 March; accepted 8 June 1998.

1. Alcock, C. *et al.* The MACHO project Large Magellanic Cloud microlensing results from the first two years and the nature of the Galactic dark halo. *Astrophys. J.* **486**, 697–726 (1997).
2. Alcock, C. *et al.* Possible gravitational microlensing of a star in the Large Magellanic Cloud. *Nature* **365**, 621–623 (1993).
3. Aubourg, E. *et al.* Evidence for gravitational microlensing by dark objects in the galactic halo. *Nature* **365**, 623–625 (1993).
4. Larson, R. B. Dark matter—dead stars? *Comments Astrophys.* **11**, 273–282 (1987).
5. Charlot, S. & Silk, J. Signatures of white dwarf galaxy halos. *Astrophys. J.* **445**, 124–132 (1995).
6. Adams, F. & Laughlin, G. Implications of white dwarf galactic halos. *Astrophys. J.* **468**, 586–597 (1996).
7. Chabrier, G., Segretain, L. & Mera, D. Contribution of brown dwarfs and white dwarfs to recent microlensing observations and the halo mass budget. *Astrophys. J.* **468**, L21–L24 (1996).
8. Graff, D. S., Laughlin, G. & Freese, K. MACHOs, white dwarfs and the age of the universe. *Astrophys. J.* **499**, 7–19 (1998).
9. D'Antona, F. & Mazzitelli, I. Cooling of white dwarfs. *Annu. Rev. Astron. Astrophys.* **28**, 139–181 (1990).
10. Bergeron, P., Saumon, D. & Wesemael, F. New model atmospheres for very cool white dwarfs with mixed H/He and pure He compositions. *Astrophys. J.* **443**, 764–779 (1995).
11. Bergeron, P., Ruiz, M. T. & Leggett, S. K. The chemical evolution of cool white dwarfs and the age of the local galactic disk. *Astrophys. J. Suppl.* **108**, 339–387 (1997).
12. Hansen, B. M. S. & Phinney, E. S. Stellar forensics I: cooling curves. *Mon. Not. R. Astron. Soc.* **294**, 557–568 (1998).

13. Mihalas, D. *Stellar Atmospheres* (Freeman, San Francisco, 1970).
14. Saumon, D., Bergeron, P., Lunine, J. I., Hubbard, W. B. & Burrows, A. Cool zero-metallicity stellar atmospheres. *Astrophys. J.* **424**, 333–344 (1994).
15. Wood, M. A. in *White Dwarfs* (eds Koester, D. & Werner, K.) 41 (Springer, Heidelberg, 1995).
16. Salaris, M. *et al.* The cooling of CO white dwarfs: influence of the internal chemical distribution. *Astrophys. J.* **486**, 413–419 (1997).
17. Liebert, J., Dahn, C. D. & Monet, D. G. The luminosity function of white dwarfs. *Astrophys. J.* **332**, 891–909 (1988).
18. Flynn, C., Gould, A. & Bahcall, J. N. Hubble Deep Field constraint on baryonic dark matter. *Astrophys. J.* **466**, L55–L58 (1996).
19. Elson, R. A. W., Santiago, B. X. & Gilmore, G. F. Halo stars, starbursts, and distant globular clusters: A survey of unresolved objects in the Hubble Deep Field. *New Astron.* **1**, 1–16 (1996).
20. Mendez, R. A., Minniti, D., de Marchi, G., Baker, A. & Couch, W. J. Starcounts in the Hubble Deep Field: constraining galactic structure models. *Mon. Not. R. Astron. Soc.* **283**, 666–672 (1996).
21. Oswalt, T. D., Smith, J. A., Wood, M. A. & Hintzen, P. A lower limit of 9.5 Gyr on the age of the galactic disk from the oldest white dwarf stars. *Nature* **382**, 692–694 (1996).
22. Clemens, J. C. The pulsation properties of the DA white dwarf variables. *Baltic Astron.* **2**, 407–434 (1993).
23. Iben, I. A. & MacDonald, J. The effects of diffusion due to gravity and due to composition gradients on the rate of hydrogen burning in a cooling degenerate dwarf. I. The case of a thick helium buffer layer. *Astrophys. J.* **296**, 540–553 (1985).
24. Paczynski, B. Evolution of single stars. VI. Model nuclei of planetary nebulae. *Acta Astron.* **21**, 417–435 (1971).
25. Richer, H. B. *et al.* White dwarfs in globular clusters: Hubble Space Telescope observations of M4. *Astrophys. J.* **484**, 741–760 (1997).
26. Bahcall, J. N., Schmidt, M. & Soneira, R. M. The galactic spheroid. *Astrophys. J.* **265**, 730–747 (1983).
27. Chaboyer, B., DeMarque, P., Kernan, P. J. & Krauss, L. M. The age of globular clusters in the light of Hipparcos: resolving the age problem? *Astrophys. J.* **494**, 96–110 (1998).
28. Holtzman, J. A. *et al.* The photometric performance & calibration of WFPC2. *Publ. Astron. Soc. Pacif.* **107**, 1065–1093 (1995).

Acknowledgements. I thank the Aspen Center for Physics for hospitality, and M. Wood, G. Chabrier, E. Garcia-Berro and M. Hernanz for comments.

Correspondence and requests for materials should be addressed to the author (e-mail: hansen@cita.utoronto.ca).

A symmetrically pulsed jet of gas from an invisible protostar in Orion

Hans Zinnecker*, Mark J. McCaughrean* & John T. Rayner†

* *Astrophysikalisches Institut Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany*

† *Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96882, USA*

Young stars are thought to accumulate most of their mass through an accretion disk, which channels the gas and dust of a collapsing cloud onto the central protostellar object¹. The rotational and magnetic forces in the star–disk system often produce high-velocity jets of outflowing gas^{2–6}. These jets can in principle be used to study the accretion and ejection history of the system, which is hidden from direct view by the dust and dense gas of the parent cloud. But the structures of these jets are often too complex to determine which features arise at the source and which are the result of subsequent interactions with the surrounding gas. Here we present infrared observations of a very young jet driven by an invisible protostar in the vicinity of the Horsehead nebula in Orion. These observations reveal a sequence of geyser-like eruptions occurring at quasi-regular intervals and with near-perfect mirror symmetry either side of the source. This symmetry is strong evidence that such features must be associated with the formation of the jet, probably related to recurrent or even chaotic instabilities in the accretion disk.

The infrared and millimetre continuum source IRAS05413–0104 is a cold (~ 30 K), low-luminosity (~ 15 times the solar luminosity) young protostar^{7–9}. The molecular cloud core in which it is embedded has a typical gas density of $\sim 10^4$ cm^{−3}, but the exceptionally low turbulent velocity dispersion¹⁰ and presence of strong water maser emission^{7,11} at 1.3 cm wavelength single it out as unusual compared to others in Orion^{7,8,10}.

Initial near-infrared observations of the core, using an infrared

camera with a small field of view, showed a pair of bright sources interpreted as a young binary stellar system¹². However, subsequent observations showed them to be the innermost of a linear chain of knots at the base of a large jet of shocked molecular hydrogen (H₂) gas. Figure 1 shows the jet, now named HH212—for Herbig–Haro object 212 (ref. 13)—in the rotational–vibrational $v = 1-0$ S(1) emission line of H₂ at a wavelength of 2.122 μ m. Strong H₂ emission is often seen from young jets¹⁴, although it does not delineate the bulk jet fluid, but rather shows radiative cooling from shocks where the fluid interacts with itself or with the ambient molecular cloud. Fortunately, many bright H₂ lines are found at near-infrared wavelengths where the extinction is relatively low, thus allowing us to study young jets still embedded in their dusty natal material. Spectroscopy (Fig. 2) demonstrates that the 1.62–2.42 μ m emission from HH212 is almost exclusively in lines of shocked H₂ at an excitation temperature of $\sim 2,400$ K, with just the tips of the brightest knots also exhibiting higher excitation [Fe II] emission at 1.644 μ m. No evidence is seen for the Br γ line of ionized hydrogen at 2.166 μ m, although for a typical ionized jet, the estimated Br γ surface brightness would be factors of ~ 10 – 100 below the current detection limit. Thus, nothing definitive can be said about the ionization fraction in HH212 based on the present data, although because it is one of the few jet sources not showing 3.6 cm radio continuum emission from ionized gas¹⁵, the jet fluid may be predominantly molecular.

High-resolution spectroscopy of the two centre knots (Fig. 2) shows only a small velocity shift of ~ 9 km s^{−1} between them, implying that the flow lies within $\sim 1^\circ$ of the plane of the sky, assuming a velocity of 200 km s^{−1} typical for young stellar jets^{16,17}. Thus any circumstellar disk around the central source must be seen nearly edge-on. Such a geometry, combined with the extended dense envelope normally associated with a protostar, explains the invisibility of the central source at these short wavelengths. The disk and envelope—as well as outflowing CO and SiO molecules in the jet—are seen directly in millimetre interferometry observations (K. B. Marvel, M.J.McC and A. I. Sargent, manuscript in preparation), while at near-infrared continuum wavelengths, two faint broad parabolic nebulae appear to delineate the concave surfaces of the disk/envelope illuminated by light reflected from the central source, as in the HH3 disk/jet system¹⁸.

HH212 covers a total length of ~ 240 arcsec, or ~ 0.5 pc ($\sim 10^5$ AU) at the 400-pc distance to the parent molecular cloud, and shows a remarkable bipolar symmetry centred on the driving source (Fig. 1b). A series of knots is seen near the source, mostly in pairs positioned at equal distances on the two sides (NK1–7 and SK1–7). These inner knots are all resolved, with intrinsic diameters of ~ 250 – 500 AU; they are wider transverse to the jet axis than parallel to it and convex in the direction of the outflow, and thus are probably small bow shocks in the flow. Further out, where the jet appears to leave the confines of the gas core¹⁰, the first of the chief bow shocks are seen; these are symmetric even to the extent that both are split into two components (NB1/2 and SB1/2). Beyond these, a pair of large, fully developed bow shocks is seen (NB3 and SB3), until finally the gross symmetry is broken by another large bow on the south side (SB4) which has no northern counterpart (as searched for in wider-field images than shown here). The pairs of features are generally equidistant from the centre to within 5–15%, with two distinct scale lengths involved (Table 1): the spacing of the brighter inner knots is $\sim 2,000$ AU, whereas the larger bow shock systems are at $\sim 16,000$ – $18,000$ AU intervals. Again assuming a constant jet outflow velocity of 200 km s^{−1}, these distances correspond to timescales of ~ 50 and 500 years, respectively.

HH212 is by far the most symmetric jet found to date, compared to systems such as the rather complex HH47¹⁹, or even the more regular HH111, where some symmetry is seen^{20,21}. In HH212, even the deviations from perfectly regular spacing are also symmetric about the centre. If symmetry is an inherent part of the jet-

formation process, it must be erased in most sources. This may be due to large-scale non-uniformity in the surroundings, either density inhomogeneities in the parent molecular cloud, or intense radiation from nearby massive stars. Alternatively, some older bipolar jets exhibit velocity differences of a factor of two or so

between the two sides, indicating that anisotropies near the source, such as pressure gradients in the collimation zone, may also play a role in symmetry-breaking¹⁶.

The high degree of spatial symmetry in HH212 appears to rule out the possibility that the features are imposed on an originally

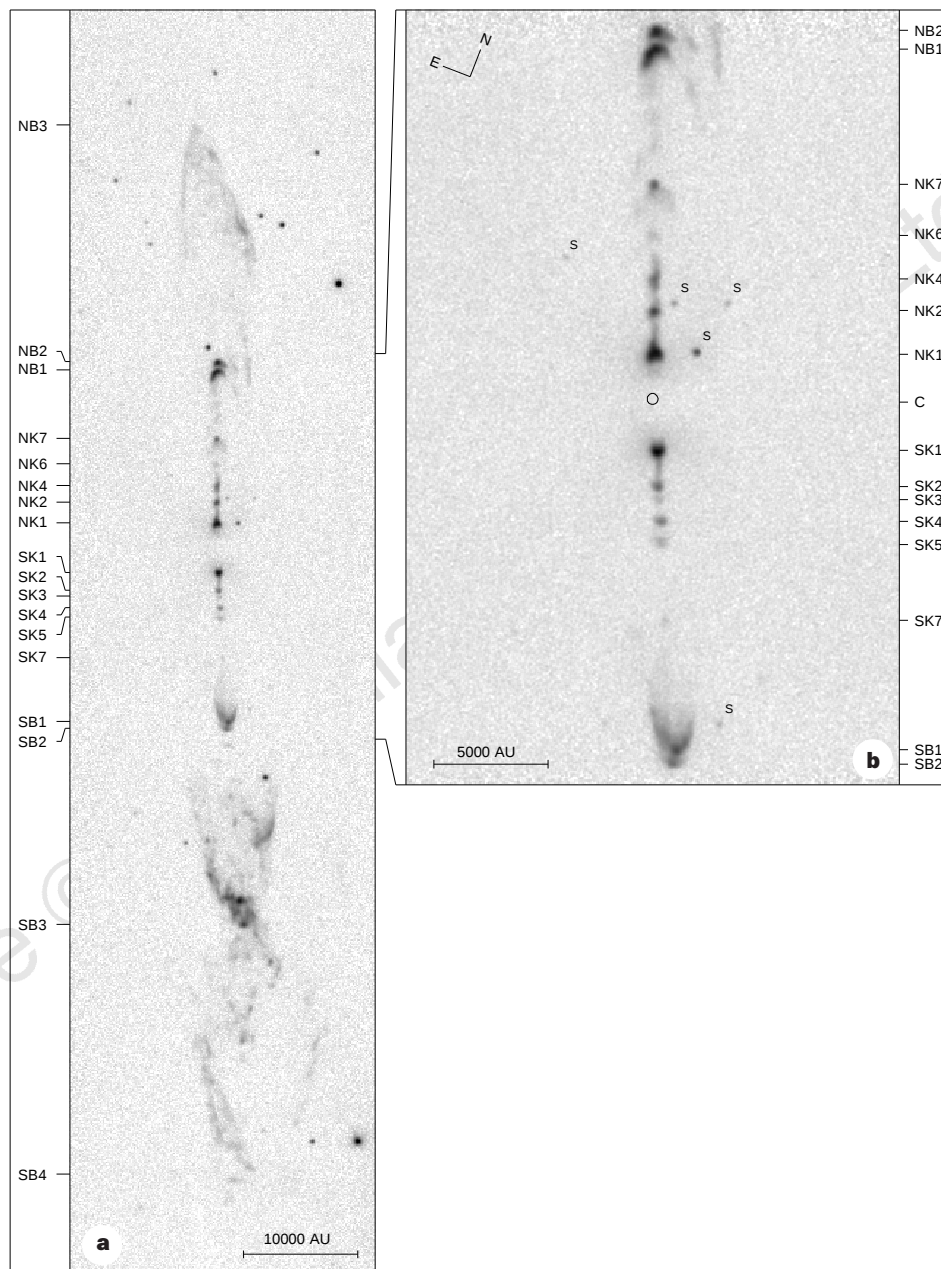


Figure 1 Molecular hydrogen images of HH212. Initial discovery observations were made on 18 December 1993 at the 3-m NASA Infrared Telescope Facility on Mauna Kea, Hawaii. The more sensitive observations presented here were taken on the Calar Alto 3.5-m telescope on 18 November 1994, using the MAGIC 256 × 256 pixel HgCdTe array camera³⁵, through a 1% bandwidth filter centred on the H₂ $\nu = 1-0$ S(1) line at 2.122 μ m. The image scale was 0.322 arcsec per pixel and field-of-view 82 × 82 arcsec. The seeing was 0.7 arcsec. A series of 1-min integrations were taken at positions along the jet, giving a total of 12 min per pixel over the full length. Images totalling a further 15 min per pixel were obtained for the central part of the jet (NB2 to SB2) the following night in better seeing (0.6 arcsec). Data reduction involved subtraction of nearby blank sky images and flat-fielding with tungsten-illuminated dome images¹⁴. A mosaic covers the whole jet (**a**: 68 × 280 arcsec) with a uniform 12 min per pixel while in the centre the many overlapping images were carefully aligned to sub-pixel accuracy giving a deep (27 min per pixel), high-resolution (0.6 arcsec) view (**b**: 55 × 86 arcsec) of the

inner knots and bow shocks. Unrelated field stars are marked 'S' in **b**. Continuum images were not subtracted, as our spectra (Fig. 2) establish that the jet is emitting only in H₂. For flux calibration, a bright standard star was observed at similar airmasses (1.4–1.7) to HH212. The resulting 3 σ detection limits are 5.5×10^{-19} and 3.7×10^{-19} W m⁻² arcsec⁻² for the full image (**a**) and central image (**b**), respectively. The total observed 2.122- μ m flux from HH212 is $(1.07 \pm 0.05) \times 10^{-15}$ W m⁻², equivalent to $5.3 \times 10^{-3} L_{\odot}$ at a distance of 400 pc. Spectra of the jet (Fig. 2), along with broad-band near-infrared observations of it and surrounding field stars, yield at estimated extinction of $A_V \approx 20$ mag ($A_{2.122 \mu m} \approx 2.8$ mag) towards the central knots, tapering off to $A_V \approx 2$ mag ($A_{2.122 \mu m} \approx 0.28$ mag) at the outer bow shocks. Correcting for this non-uniform extinction yields an intrinsic 2.122- μ m flux of 3.63×10^{-15} W m⁻² or $0.018 L_{\odot}$. For the derived jet excitation parameters (Fig. 2), the $\nu = 1-0$ S(1) line accounts for 6.5% of the total H₂ luminosity³⁶, yielding a total inferred H₂ luminosity for HH212 of $0.28 L_{\odot}$. Cumulative errors are of the order of $\pm 50\%$.

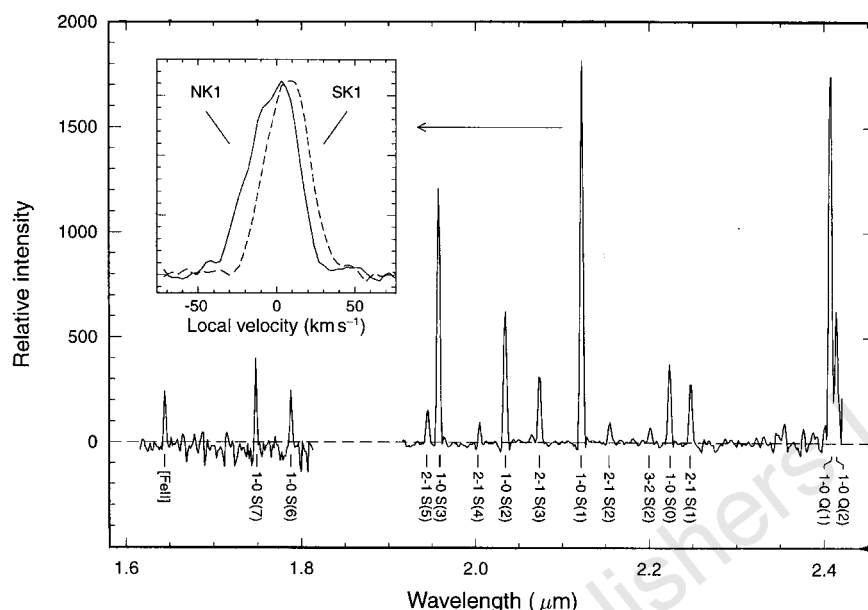


Figure 2 Infrared spectroscopy of HH212. The main plot shows a medium-resolution (resolution $R \approx 500$) spectrum covering 1.62–2.42 μm taken on the University of Hawaii 2.1-m telescope on Mauna Kea on 20 December 1994 using the cross-dispersed KSPEC spectrometer³⁷. The spectrum is a mean of separate spectra taken of the brightest inner knots, NK1 and SK1. Data acquisition and reduction was standard, including sky subtraction, flat-fielding, and compensation for telluric absorption features. The cumulative on-source integration time was 18 min. In the 2- μm window, only H_2 line emission is seen from the jet: there is no detectable continuum component or any trace of hydrogen recombination emission at the 2.166- μm $\text{Br}\gamma$ line. Emission is seen from the higher excitation tracer line of [Fe II] at 1.644 μm . Standard analysis of the H_2 line ratios^{36,38,39} shows shocked (rather than fluorescent) excitation at a temperature

2,400 ± 80 K, a hydrogen molecule *ortho-para* ratio of 3.1 ± 0.5 , a shock speed of 10 km s^{-1} (for a J-shock) or 45 km s^{-1} (for a magnetized C-shock), and an extinction towards the inner jet of $A_V \approx 20$ mag. Inset, velocity-resolved spectroscopy ($R \approx 21,000$, 14 km s^{-1}) of the $v = 1-0 \text{ S}(1)$ line for NK1 and SK1 obtained using the CSHELL echelle spectrometer⁴⁰ on the 3.0-m NASA IRTF on Mauna Kea on 19 January 1995. Standard data acquisition and reduction techniques were again employed. The total on-source integration time was 27 min. Both knots have relatively broad velocity profiles, up to 75 km s^{-1} full-width zero-intensity. The symmetric, single-peaked profiles, with knot velocity centroids that are clearly separated by $\sim 9 \text{ km s}^{-1}$, indicate that the flow lies close to the plane of the sky, with the northern lobe of HH212 approaching the Earth, and the southern lobe receding.

uniform jet via flow instabilities (for example, Kelvin–Helmholtz or magneto-hydrodynamic (MHD) pinch-mode instabilities²) as the jet propagates through, and interacts with, the ambient medium. It is much more likely that the features arise through time variability at the driving source. Indeed, the extraordinary bilateral symmetry of HH212 provides the strongest evidence yet for intrinsic ejection variability at the source, in the form of pulses in the mass loss rate, the outflow velocity, or a combination of both. We speculate that these pulses are caused by variations in the disk accretion flow (and thence jet outflow), driven by self-regulating, recurrent thermal ionization instabilities²², combined with enhanced transport due to ensuing MHD turbulence²³. Repetitive processes of this kind, related to the so-called FU Orionis phenomenon, have previously been suggested as the cause of episodic jets²⁰, and numerical models of velocity-variable jets can indeed reproduce the observed knotty shock structures^{24–26}. In HH212, there appear to be two distinct intervals (~ 50 and 500 years), although these could just be the most prominent of a hierarchy of timescales consistent with self-similar, nonlinear, chaotic behaviour characterized by a $1/f$ power spectrum²⁷ (where f is frequency). This kind of self-similar time variability is also seen, for example, in the accretion disks of low-mass X-ray binaries²⁸ and active galactic nuclei²⁹ (AGN).

A fiducial kinematic jet luminosity, L_{kin} , can be calculated for HH212 as $2 L_{\odot} (\dot{M}/10^{-6}) (v/200)^2$, for estimated mass loss rate, $\dot{M}$ (in units of $M_{\odot} \text{ yr}^{-1}$; the reference value is $\sim 10\%$ of the estimated accretion rate⁸), and velocity, v (in units of km s^{-1}), as given. L_{kin} is $\sim 10\%$ of the bolometric luminosity of the central source⁸ and ~ 10 times the jet H_2 luminosity. A kinematic luminosity of this order can easily be extracted from the rotational motion of the star and inner disk system, and converted into the Poynting flux of the MHD flow².

The H_2 line profiles (Fig. 2) allow insight into the internal shock physics of the inner knots, NK1 and SK1. Models of bow shocks viewed side-on predict that the full-width zero-intensity (FWZI) spread in radial velocities should be equal to the tangential bow velocity. However, in a high-velocity molecular shock, the maximum FWZI is equal to the H_2 dissociation speed limit, above which the molecules are destroyed and no higher velocity H_2 emission can be seen³⁰. For a normal jump or ‘J-shock’, this limit is only $\sim 25 \text{ km s}^{-1}$. However, for a magnetically cushioned ‘C-shock’, the speed limit is sensitive to the magnetic field strength, and can be considerably higher³¹. In the case of HH212, the limit of 75 km s^{-1} implied by the measured line FWZI (Fig. 2) in turn implies an Alfvén velocity of $\sim 15 \text{ km s}^{-1}$ (Fig. 2 of ref. 31), and thus a high toroidal magnetic field strength of a few milligauss near the central knots, for typical pre-shock jet densities in the range a few times 10^4 to 10^5 cm^{-3} . The magnetic field strength in the knots themselves is likely to be an order of magnitude higher, in proportion to the post-shock gas compression factor, and is perhaps observable. This is quite plausible given that strong magnetic fields have been recently detected directly via circular polarization measurements in the outflow from T Tauri S³². An alternative to the C-shock model would involve molecules reforming behind a dissociative J-shock and then emitting. However, the canonical timescale for H_2 reformation on dust is $\sim (3 \times 10^3 \text{ yr})/(n/10^6 \text{ cm}^{-3})$ which, for this estimated post-shock density (where a gas compression of ~ 10 is assumed), exceeds the overall dynamical age of the jet. Thus reformation cannot play a significant role.

The intermittent shock structures seen in the HH212 jet emanating from a low-mass protostar are reminiscent of the knotty jets associated with galactic superluminal sources³³ and AGN³⁴. AGN

Table 1 Positions and fluxes of features in HH212

North jet feature	Position		Offset distance (AU)	Peak flux ($10^{-19} \text{ W m}^{-2} \text{ arcsec}^{-2}$)	South jet feature	Position		Offset distance (AU)	Peak flux ($10^{-19} \text{ W m}^{-2} \text{ arcsec}^{-2}$)
	RA	Dec.				RA	Dec.		
	(see below)					(see below)			
NK1	51.51	02 48.2	2,006	223	SK1	51.22	02 58.3	2,365	263
NK2	51.62	02 43.9	3,855	97	SK2	51.13	03 01.8	3,907	47
NK3					SK3	51.09	03 03.0	4,425	14
NK4	51.71	02 40.5	5,324	51	SK4	51.02	03 05.4	5,461	34
NK5					SK5	50.97	03 07.3	6,306	15
NK6	51.84	02 36.1	7,247	10	SK6				
NK7	51.96	02 30.8	9,484	55	SK7	50.71	03 15.5	9,898	7
NB1	52.31	02 16.6	15,530	187	SB1	50.30	03 28.2	15,565	87
NB2	52.36	02 14.9	16,269	276	SB2	50.29	03 29.7	16,142	45
NB3	53.94	01 28.0	37,259	10–28	SB3	49.00	04 08.8	33,561	25–216
NB4					SB4	47.36	04 58.4	55,720	5–15

Column headings have meanings as follows. 'North jet feature' and 'South jet feature' correspond to the features marked in Fig. 1 on the NNE and SSW sides of HH212, respectively. 'Position', the coordinate of the feature, equinox J2000.0. The right ascension (RA) should be prefixed with 05 h 43 min, and the declination (Dec.) with -01° . Astrometry was determined using the Digital Sky Survey and stars from the Hubble Space Telescope Guide Star Catalog (version 1.2), creating a local reference frame for a wide-field broad-band K image, which was then transferred to the H₂ image. The resulting astrometry in the GSC v 1.2 frame has a cumulative two-dimensional error of 0.29 arcsec r.m.s. 'Offset distance', distance of the feature from the millimetre-continuum source position at 05 h 43 min 51.39 s $-01^\circ 02' 9''$ (J2000.0), as marked with a circle in Fig. 1. The distance (astronomical units) assumes HH212 to be 400 parsec from the Earth. 'Peak flux', peak observed (that is, not corrected for extinction) surface brightness of feature in the $v = 1-0 \text{ S}(1) \text{ H}_2$ line.

jets are on a vastly larger scale, are ionized rather than molecular, underdense relative to the ambient medium rather than overdense, and emit predominantly continuum rather than line radiation. Nevertheless, perhaps there is a common underlying physical process in all astrophysical jets—self-regulating instabilities in magnetized accretion disks. In this sense, HH212 may be a 'standard' jet, and its simplicity suggests that detailed future studies may significantly improve our understanding of the accretion/ejection connection both in young stellar objects and AGN. □

Received 23 February; accepted 28 May 1998.

- Beckwith, S. V. W. & Sargent, A. I. Circumstellar disks and the search for neighbouring planetary systems. *Nature* **383**, 139–144 (1996).
- Camenzind, M. in *Herbig-Haro Flows and the Birth of Low Mass Stars* (eds Reipurth, B. & Bertout, C.) 241–258 (Proc. IAU Symp. 182, Kluwer, Dordrecht, 1997).
- Shu, F. H. & Shang, H. in *Herbig-Haro Flows and the Birth of Low Mass Stars* (eds Reipurth, B. & Bertout, C.) 225–239 (Proc. IAU Symp. 182, Kluwer, Dordrecht, 1997).
- Oued, R., Pudritz, R. E. & Stone, J. M. Episodic jets from black holes and protostars. *Nature* **385**, 409–414 (1997).
- Edwards, S., Ray, T. & Mundt, R. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. L.) 567–602 (Univ. Arizona Press, Tucson, 1993).
- Bachiller, R. Bipolar molecular outflows from young stars and protostars. *Annu. Rev. Astron. Astrophys.* **34**, 111–154 (1996).
- Wouterloot, J. G. A. & Walmsley, C. M. H₂O masers associated with IRAS sources in regions of star formation. *Astron. Astrophys.* **168**, 237–247 (1986).
- Zinnecker, H., Bastien, P., Arcoragi, J.-P. & Yorke, H. W. Submillimeter dust continuum observations of three low luminosity protostellar IRAS sources. *Astron. Astrophys.* **265**, 726–732 (1992).
- Chini, R. *et al.* Cold dust around Herbig-Haro energy sources: morphology and new protostellar candidates. *Astron. Astrophys.* **325**, 542–550 (1997).
- Wouterloot, J. G. A., Henkel, C. & Walmsley, C. M. CO observations of IRAS sources in Orion and Cepheus. *Astron. Astrophys.* **215**, 131–146 (1989).
- Clausen, M. J., Marvel, K. B., Wootten, H. A. & Wilking, B. A. in *Herbig-Haro Flows and the Birth of Low Mass Stars* (eds Reipurth, B. & Bertout, C.) 515–524 (Proc. IAU Symp. 182, Kluwer, Dordrecht, 1997).
- Zinnecker, H. in *Low-mass Star Formation and Pre-main Sequence Objects* (ed. Reipurth, B.) 447–469 (Proc. ESO Workshop, European Southern Observatory, Garching, 1989).
- Reipurth, B. A general catalog of Herbig-Haro objects. Available by anonymous ftp, directory pub/Catalogs/Herbig-Haro from (ftp.eso.org) (1994).
- McCaughrean, M. J., Rayner, J. T. & Zinnecker, H. Discovery of a molecular hydrogen jet near IC348. *Astrophys. J.* **436**, L189–L192 (1994).
- Rodríguez, L. F. & Reipurth, B. VLA detection of the exciting sources of HH24, HH114, and HH199. *Rev. Mex. Astron. Astrofis.* **32**, 27–33 (1996).
- Hirth, G., Mundt, R., Solf, J. & Ray, T. P. Asymmetries in bipolar jets from young stars. *Astrophys. J.* **427**, L99–L102 (1994).
- Fendt, C. & Camenzind, M. On collimated stellar jet magnetospheres. II. Dynamical structure of collimating wind flows. *Astron. Astrophys.* **313**, 591–604 (1996).
- Burrows, C. J. *et al.* Hubble Space Telescope observations of the disk and jet of HH30. *Astrophys. J.* **473**, 437–451 (1996).
- Heathcote, S. *et al.* Hubble Space Telescope observations of the HH47 jet: narrowband images. *Astron. J.* **112**, 1141–1168 (1996).
- Reipurth, B. The HH111 jet and multiple outflow episodes from young stars. *Nature* **340**, 42–45 (1989).
- Gredel, R. & Reipurth, B. An infrared counter-flow in the HH111 jet complex. *Astron. Astrophys.* **289**, L19–L22 (1994).
- Bell, K. R. & Lin, D. N. C. Using FU Orionis outbursts to constrain self-regulated protostellar disk models. *Astrophys. J.* **427**, 987–1004 (1994).
- Balbus, S. A. & Hawley, J. F. Instability, turbulence, and enhanced transport in accretion disks. *Rev. Mod. Phys.* **70**, 1–54 (1998).
- Stone, J. M. & Norman, M. L. Numerical simulations of protostellar jets with nonequilibrium cooling. II. Models of pulsed jets. *Astrophys. J.* **413**, 210–220 (1993).
- Raga, A. C. & Kofman, L. Knots in stellar jets from time-dependent sources. *Astrophys. J.* **386**, 222–228 (1992).
- Suttnner, G., Smith, M. D., Yorke, H. W. & Zinnecker, H. Multi-dimensional numerical simulations of molecular jets. *Astron. Astrophys.* **318**, 595–607 (1997).

- Reipurth, B., Bally, J. & Devine, D. Giant Herbig-Haro flows. *Astron. J.* **114**, 2708–2735 (1997).
- van der Klis, M. Quasi-periodic oscillations and noise in low-mass X-ray binaries. *Annu. Rev. Astron. Astrophys.* **27**, 517–553 (1989).
- McHardy, I. & Czerny, B. Fractal X-ray time variability and spectral invariance of the Seyfert galaxy NGC5506. *Nature* **325**, 696–698 (1987).
- Smith, M. D. & Brand, P. W. J. L. H₂ profiles of C-type bow shocks. *Mon. Not. R. Astron. Soc.* **245**, 108–118 (1990).
- Smith, M. D., Brand, P. W. J. L. & Moorhouse, A. Shock absorbers in bipolar outflows. *Mon. Not. R. Astron. Soc.* **248**, 730–740 (1991).
- Ray, T. P. *et al.* Large-scale magnetic fields in the outflow from the young stellar object T Tauri S. *Nature* **385**, 415–417 (1997).
- Mirabel, I. F. & Rodríguez, L. F. A superluminal source in the galaxy. *Nature* **371**, 46–48 (1994).
- Zensus, J. A. Parsec-scale jets in extragalactic radio sources. *Annu. Rev. Astron. Astrophys.* **35**, 605–636 (1997).
- Herbst, T. M. *et al.* in *Infrared Detectors and Instrumentation* (ed. Fowler, A. M.) *Proc. SPIE* **1946**, 605–609 (1993).
- Smith, M. D. Predictions for JHK photometry of molecular shocks. *Astron. Astrophys.* **296**, 789–796 (1995).
- Hodapp, K.-W., Hora, J. L., Erwin, E. & Young, T. KSPC—a near infrared cross-dispersed spectrograph. *Publ. Astron. Soc. Pacif.* **106**, 87–93 (1994).
- Smith, M. D., Davis, C. J. & Liou, A. The ortho and para fractions of molecular hydrogen in protostellar outflows and Herbig-Haro objects. *Astron. Astrophys.* **327**, 1206–1214 (1997).
- Herbst, T. M. *et al.* A near-infrared spectral imaging study of T Tauri. *Astron. J.* **111**, 2403–2414 (1996).
- Greene, T. P., Tokunaga, A. T., Toomey, D. W. & Carr, J. S. in *Infrared Detectors and Instrumentation* (ed. Fowler, A. M.) *Proc. SPIE* **1946**, 313–324 (1993).

Acknowledgements. We thank S. Balbus, J. Bally, R. Blandford, W. Brandner, M. Camenzind, W. Dent, C. Fendt, R. Gredel, G. Hasinger, L. Kofman, R. Mundt, M. Norman, M. Rees, B. Reipurth, M. Smith, F. Shu, J. Stone, C. Terquem, M. Walmsley and H. Yorke for discussions on the theory and observations of astrophysical jets. We also thank C. Ishida for reducing the KSPC data. H.Z. and M.J.M. were visiting astronomers at the NASA IRTF using the facility camera NSFACAM, with which HH212 was originally discovered.

Correspondence and requests for materials should be addressed to H.Z. (e-mail: hzinnecker@aip.de).

Experimental evidence for microscopic chaos

P. Gaspard*, M. E. Briggs†, M. K. Francis‡, J. V. Sengers‡, R. W. Gammon‡, J. R. Dorfman‡ & R. V. Calabrese‡

* Université Libre de Bruxelles, CP 231-Campus Plaine, Boulevard du Triomphe, B-1050 Brussels, Belgium

† University of Utah, Salt Lake City, Utah 84112, USA

‡ University of Maryland, College Park, Maryland 20742, USA

Many macroscopic dynamical phenomena, for example in hydrodynamics and oscillatory chemical reactions, have been observed to display erratic or random time evolution, in spite of the deterministic character of their dynamics—a phenomenon known as macroscopic chaos^{1–5}. On the other hand, it has been long supposed that the existence of chaotic behaviour in the microscopic motions of atoms and molecules in fluids or solids is responsible for their equilibrium and non-equilibrium properties. But this hypothesis of microscopic chaos has never been

verified experimentally. Chaotic behaviour of a system is characterized by the existence of positive Lyapunov exponents, which determine the rate of exponential separation of very close trajectories in the phase space of the system⁶. Positive Lyapunov exponents indicate that the microscopic dynamics of the system are very sensitive to its initial state, which, in turn, indicates that the dynamics are chaotic; a small change in initial conditions will lead to a large change in the microscopic motion. Here we report direct experimental evidence for microscopic chaos in fluid systems, obtained by the observation of brownian motion of a colloidal particle suspended in water. We find a positive lower bound on the sum of positive Lyapunov exponents of the system composed of the brownian particle and the surrounding fluid.

Numerical and theoretical work suggests that the motion of many-particle systems is chaotic, but the timescale of this microscopic chaos is much smaller than that for macroscopic chaos^{7–10}. This microscopic motion displays considerable dynamical instability due to the defocusing character of the collisions between the fluid particles, leading to a sensitivity to initial conditions on a timescale of the order of the mean time between collisions. In turn, the instability of motion, which is characterized by the Lyapunov exponents, causes the particles to evolve along erratic trajectories (see Fig. 1).

More than 50 years ago, Krylov pointed out¹¹ that the microscopic dynamical instability has a role to play in the relaxation of fluids to equilibrium. This role has been much strengthened by recent theoretical work which has established explicit relationships

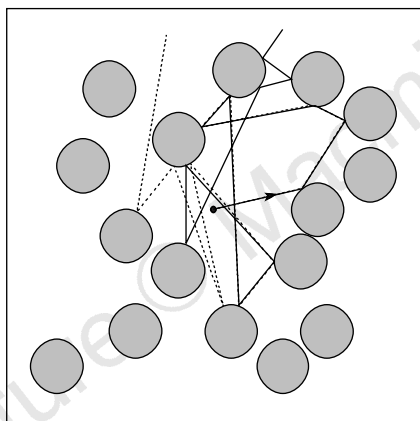


Figure 1 Dynamical instability in a Lorentz gas. Illustration of the mechanism by which an early dynamical instability can generate temporal randomness due to the collisions between the particles of a fluid. Shown are two trajectories in a random Lorentz gas in which a particle undergoes elastic collisions with a random planar configuration of immobile hard disks²⁹. The initial conditions are very close to each other, but both trajectories separate after a few collisions and display erratic motion typical of brownian motion. In a gas at standard temperature and pressure, the Lyapunov exponents are of the order of the inverse of the mean free time $\lambda_f \approx 10^{10}$ digits s^{-1} . Because perturbations can be considered on each particle of the system, the number of positive Lyapunov exponents is equal to $3N - 1$ for a system of N particles in which energy is conserved.

between microscopic chaos and transport phenomena such as diffusion, viscosity or heat conduction, as well as in the positivity of entropy production^{10,12–17}. In this context, Gallavotti and Cohen have proposed¹⁸ a 'chaotic hypothesis' which supposes that many important properties of the many-particle systems of statistical mechanics can be predicted by treating the systems as if they are, essentially, Anosov systems—smooth chaotic systems¹⁰ with microscopic behaviour characterized by positive Lyapunov exponents.

In view of these studies, we have set out to look for direct, experimental evidence of the existence of this hypothesized microscopic chaos^{19,20}. We can test for chaos by searching for recurrences of patterns in the trajectory of the system during a long series of repeated measurements. The absence of more recurrences than statistically expected is a test of randomness invented by Chaitin²¹ and Kolmogorov²². According to this test, the probability of observing a trajectory pattern should decrease exponentially as the length of the pattern increases, if the motion is chaotic. The rate of exponential decrease of the pattern probabilities is known as the entropy per unit time and can be interpreted as the rate of production of information on the trajectory of the system^{6,23,24}. The positivity of the entropy per unit time is thus an evidence for randomness in the time series.

Because randomness is generated in an Anosov-like system by the underlying dynamical instability, the pattern probabilities cannot decrease at a rate larger than provided by all the positive Lyapunov exponents. Therefore, the entropy per unit time is always smaller than the sum of the positive Lyapunov exponents. For such systems,

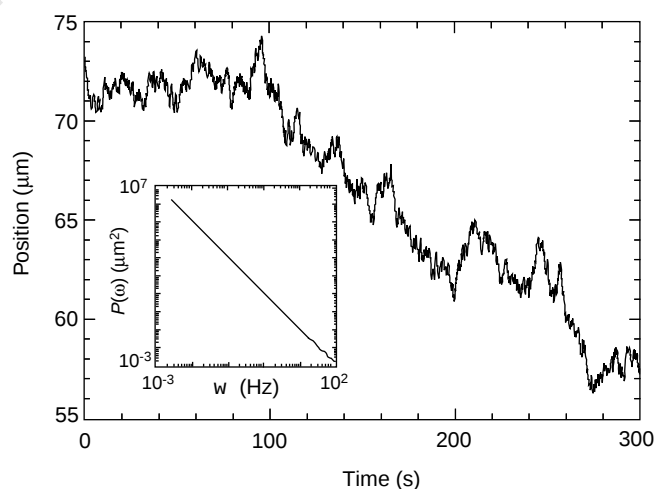


Figure 2 Brownian trajectory and power spectrum. Portion of the brownian trajectory recorded in the present experiment and the corresponding power spectrum $P(\omega)$ (inset). The complete time series contains 145,612 positions over a total time interval of $\sim 2,430$ s. The brownian particle has a diameter of $2.5 \mu m$ and moves in suspension in deionized water at $22^\circ C$, between a polished silicon wafer and glass coverslip, with a cell height of $50 \mu m$ and a seal of vacuum grease. The sampling time is $\tau = 1/60$ s. The smallest resolution allowed by the microscope and the video camera is $\epsilon_{min} = 25$ nm. The particle tracking procedure is essentially the same as that described by Grier and Murray²⁰. An estimate of the resolution of the experiment is the power spectrum measured on a fixed particle of the same type. Such a measurement gives a flat spectrum one decade below the lowest power shown on this figure. A power spectrum as $P(\omega) \sim \omega^{-2}$ over more than four frequency decades shows that the motion is of brownian character with a diffusion coefficient $D \approx 0.124 \mu m^2 s^{-1}$. The reduced diffusion coefficient occurs because a particle of this size sediments and thus experiences a wall-drag effect. The diffusion coefficient is $\sim 30\%$ below that expected for free diffusion, which is consistent with other work³¹. We also studied $1\text{-}\mu m$ particles which do not sediment, and found the expected free diffusion coefficient, and the same dynamic entropy behaviour as shown by the large particles. However, the larger particles can be tracked for a longer time, and provide a larger scaling range. We report only the results for the $2.5\text{-}\mu m$ particle here.

the sum of positive Lyapunov exponents is known to be equal to the Kolmogorov–Sinai entropy which is the maximum possible rate of exponential decrease of the pattern probabilities^{6,20,25}. This quantitative connection between instability and chaos is an essential result, as it provides a way to test the dynamical instability of motion by measuring the temporal randomness of a natural process.

The experimental test reported here of the hypothesis of microscopic chaos is based on the observation that many processes taking place in fluids display a high degree of temporal randomness; here we focus on the erratic brownian motion of a colloidal particle in a liquid. Our purpose is to measure the entropy per unit time of such a brownian trajectory by the method mentioned above, which should provide a lower bound on the sum of positive Lyapunov exponents. Therefore, the measurement of a positive entropy per unit time would be evidence for the positivity of some Lyapunov exponents and, hence, for microscopic chaos. This is, to our knowledge, the first time such direct evidence of microscopic chaos has been observed in a laboratory experiment.

We have carried out an experiment that records a long time series of the position of a brownian particle as a function of time (Fig. 2). We have calculated the probabilities $P(X_m; n, \epsilon, \tau)$ for the trajectory to remain within a distance ϵ of M reference trajectories X_m , made of n successive positions of the brownian particle observed at time intervals τ . These reference trajectories define the patterns, the recurrences of which are searched for in the time series. According to a numerical procedure proposed by Grassberger, Procaccia and others,^{23,24} the entropy per unit time can be estimated by the linear growth of the mean ‘pattern entropy’ defined as:

$$K(n, \epsilon, \tau) = -\frac{1}{M} \sum_{m=1}^M \log_{10}[P(X_m; n, \epsilon, \tau)] \quad (1)$$

This quantity is depicted in Fig. 3 as a function of the number of measurement times $n\tau$ for different values of the distance ϵ . Figure 3 shows linear growth of the mean pattern entropy with the number of times $n\tau$ and, hence, the positivity of the entropy per unit time $h(\epsilon, \tau)$ which can be evaluated from the slopes of the lines in Fig. 3 for the different values of the distance ϵ (refs 23, 24). For brownian motion, the entropy per unit time is a function of both the distance ϵ and the sampling time τ , as seen in Fig. 4. The different curves form an envelope which scales in a way directly related to the diffusive

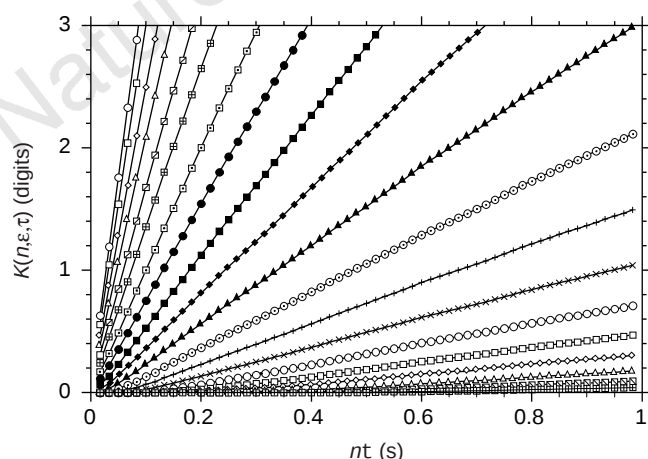


Figure 3 The mean pattern entropy as a function of time. Shown is $K(n, \epsilon, \tau)$ defined by equation (1) versus $n\tau$ for the sampling time $\tau = \Delta t = 1/60$ s and for different values of the distance $\epsilon = (1.2)^n \times 0.03 \mu\text{m}$ with $n = 1, \dots, 20$. The distance used in the calculation is defined by taking the maximum among the distances $|X(t) - X(0) - X_m(t) + X_m(0)|$ for the times $t = 0, \tau, 2\tau, \dots, (n-1)\tau$. The larger slopes correspond to the smaller values of ϵ . The linear growth persists up to a maximum value of the mean pattern entropy given by the total length of the time series: $K_{\text{max}} = \log_{10}(145, 612)$.

character of brownian motion:

$$h(\epsilon, \tau) \leq A \frac{D}{\epsilon^2} \quad (2)$$

where D is the diffusion coefficient and $A \approx 1.34$ digits (unit of base-ten logarithms) is a numerical constant. This scaling behaviour has been theoretically predicted^{20,26,27} and is here experimentally confirmed, as observed in Fig. 4. The non-saturation of the entropy as the distance ϵ decreases is evidence that the microscopic chaos involves a very large number of degrees of freedom, as expected for a many-particle system.

Our quantitative measure of a positive entropy per unit time implies the existence of chaos, according to dynamical systems theory⁶. Indeed, in Anosov systems, the sum of positive Lyapunov exponents—which is equal to the Kolmogorov–Sinai entropy per unit time h_{KS} —is the *supremum* (least upper bound) over all the possible entropies per unit time like $h(\epsilon, \tau)$ (ref. 6). As a consequence, every entropy per unit time is a lower bound on the sum of the positive Lyapunov exponents λ_i :

$$h(\epsilon, \tau) \leq h_{\text{KS}} = \sum_{\lambda_i > 0} \lambda_i \quad (3)$$

which is our central argument. Taking the value of the entropy at the upper end of its scaling range at $\epsilon = 0.22 \mu\text{m}$ for $\tau = 1/60$ s in Fig. 4, we obtain the lower bound $h(\epsilon, \tau) \approx 3.2$ digits $\text{s}^{-1} \leq \sum_{\lambda_i > 0} \lambda_i$. The positivity of this lower bound is essentially due to the positive slopes in Fig. 3, and shows that the algorithm used here establishes the chaotic nature of the motion of the particles. Because brownian motion only allows the observation of a very small subset of all the microscopic degrees of freedom, the bound takes a very low value with respect to typical Lyapunov exponents. Higher values of the lower bound could certainly be obtained using higher sampling rates and higher resolution, or by observing faster processes such as Nyquist thermal noise in electric circuits. For other noises than the brownian motion, the entropy per unit time may have a different behaviour with ϵ and τ than shown in equation (2), but the inequality of equation (3) would still provide a lower bound on the positive Lyapunov exponents.

We have obtained strong experimental evidence for microscopic chaos. On the assumption that the system is deterministic, and given our knowledge of the molecular structure of the fluid, this evidence supports, in particular, the hypothesis that large systems—

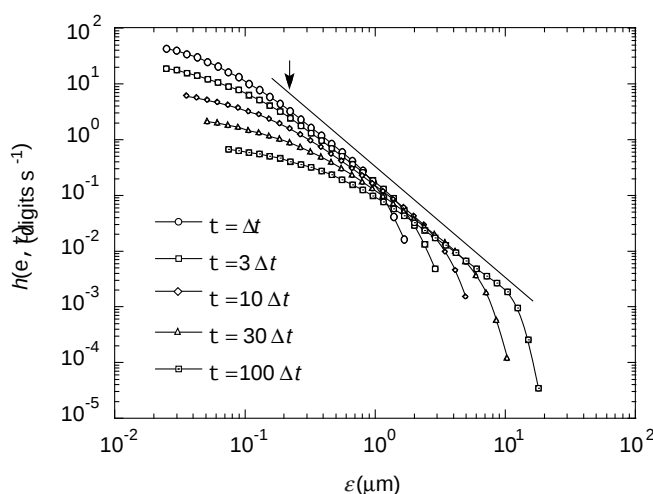


Figure 4 The entropy per unit time as a function of resolution. Shown is $h(\epsilon, \tau)$ defined by the slope of the lines in Fig. 3; that is, by the rates of linear growth of the mean pattern entropy according to the equation $K(n, \epsilon, \tau) \approx n\tau h(\epsilon, \tau)$, for different values of the sampling time $\tau = \Delta t, 3\Delta t, 10\Delta t, 30\Delta t, 100\Delta t$. The entropy per unit time is thus the mean rate of exponential decrease of the pattern probabilities. The straight line gives the slope -2 expected by the scaling law of equation (2). The vertical arrow points to the value used as the lower bound on the sum of positive Lyapunov exponents.

which may be treated by statistical mechanics—are typically chaotic. The result also supports the role of dynamical instability in nonequilibrium fluids^{12–17,28}. □

Received 29 December 1997; accepted 8 June 1998.

- Eckmann, J.-P., Oliffson Kamphorst, S., Ruelle, D. & Ciliberto, S. Lyapunov exponents from time series. *Phys. Rev. A* **34**, 4971–4979 (1986).
- Roux, J.-C., Simoyi, R. H. & Swinney, H. L. Observation of a strange attractor. *Physica D* **8**, 257–266 (1983).
- Wisdom, J., Peale, S. J. & Mignard, F. The chaotic rotation of Hyperion. *Icarus* **58**, 137–152 (1984).
- Klavetter, J. J. Rotation of Hyperion. I. Observations. *Astron. J.* **97**, 570–579 (1989).
- Laskar, J. A numerical experiment on the chaotic behaviour of the Solar System. *Nature* **338**, 237–238 (1989).
- Eckmann, J.-P. & Ruelle, D. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985).
- Livi, R., Politi, A. & Ruffo, S. Distribution of characteristic exponents in the thermodynamic limit. *J. Phys. A* **19**, 2033–2040 (1986).
- Posch, H. A. & Hoover, W. G. Lyapunov instability of dense Lennard-Jones fluids. *Phys. Rev. A* **38**, 473–482 (1988); Equilibrium and nonequilibrium Lyapunov spectra for dense fluids and solids. *Phys. Rev. A* **39**, 2175–2188 (1989).
- Dellago, Ch., Posch, H. A. & Hoover, W. G. Lyapunov instability in a system of hard disks in equilibrium and nonequilibrium steady states. *Phys. Rev. E* **53**, 1485–1501 (1996).
- Gaspard, P. *Chaos, Scattering Theory, and Statistical Mechanics* (Cambridge Univ. Press, 1998).
- Krylov, N. Relaxation processes in statistical systems. *Nature* **153**, 709–710 (1944).
- Bunimovich, L. A. & Sinai, Ya. G. Statistical properties of Lorentz gas with periodic configuration of scatterers. *Commun. Math. Phys.* **78**, 479–497 (1981).
- Evans, D. J., Cohen, E. G. D. & Morriss, G. P. Viscosity of a simple fluid from its maximal Lyapunov exponents. *Phys. Rev. A* **42**, 5990–5997 (1990).
- Gaspard, P. & Nicolis, G. Transport properties, Lyapunov exponents, and entropy per unit time. *Phys. Rev. Lett.* **65**, 1693–1696 (1990).
- Dorfman, J. R. & Gaspard, P. Chaotic scattering theory of transport and reaction-rate coefficients. *Phys. Rev. E* **51**, 28–35 (1995).
- Ruelle, D. Positivity of entropy production in nonequilibrium statistical mechanics. *J. Stat. Phys.* **85**, 1–23 (1996).
- Chernov, N. I., Eyink, G. L., Lebowitz, J. L. & Sinai, Ya. G. Derivation of Ohm's law in a deterministic mechanical model. *Phys. Rev. Lett.* **70**, 2209–2212 (1993).
- Gallavotti, G. & Cohen, E. G. D. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.* **74**, 2694–2697 (1995); Dynamical ensembles in stationary states. *J. Stat. Phys.* **80**, 931–970 (1995).
- Gaspard, P. Can we observe microscopic chaos in the laboratory? *Adv. Chem. Phys.* **XCIX**, 369–392 (1997).
- Gaspard, P. & Wang, X.-J. Noise, chaos, and (ϵ, τ) -entropy per unit time. *Phys. Rep.* **235**, 291–345 (1993).
- Chaitin, G. J. *Algorithm Information Theory* (Cambridge Univ. Press, 1987).
- Kolmogorov, A. N. Combinatorial foundations of information theory and the calculus of probabilities. *Russ. Math. Survey* **38**, 29–40 (1983).
- Grassberger, P. & Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **28**, 2591–2593 (1983).
- Cohen, A. & Procaccia, I. Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems. *Phys. Rev. A* **31**, 1872–1882 (1985).
- Ott, E. *Chaos in Dynamical Systems* (Cambridge Univ. Press, 1993).
- Kolmogorov, A. N. On the Shannon theory of information transmission in the case of continuous signals. *IRE Trans. Information Theory* **IT-2**, 102–108 (1956).
- Berger, T. Information rates of Wiener processes. *IEEE Trans. Information Theory* **IT-16**, 134–139 (1970).
- Tél, T., Gaspard, P. & Nicolis, G. (eds) *Chaos* **8**, (Focus Issue on Chaos and Irreversibility) (1998).
- van Beijeren, H. Transport properties of stochastic Lorentz models. *Rev. Mod. Phys.* **54**, 195–234 (1982).
- Grier, D. G. & Murray, C. A. in *Structure and Dynamics of Strongly Interacting Colloids and Supramolecular Aggregates in Solution* (eds Chen, S. H., Huang, J. S. & Tartaglia, P.) 145 (NATO ASI Ser. C, 369, Kluwer, Boston, 1991).
- Schaertl, W. & Sillescu, H. Dynamics of colloidal hard spheres in thin aqueous suspension layers. *J. Colloid Interface Sci.* **155**, 313–318 (1993).

Acknowledgements. P.G. was supported by the National Fund for Scientific Research (FNRS Belgium); J.R.D. thanks T. Gilbert for helpful remarks, and the US NSF for support; M.E.B. thanks the Physics Department at the University of Utah for support. This work was supported in part by the IUAP/PAI program of the Belgium Government and by the Banque Nationale de Belgique.

Correspondence and requests for materials should be addressed to P.G. (e-mail: gaspard@ulb.ac.be).

Controlling local disorder in self-assembled monolayers by patterning the topography of their metallic supports

Joanna Aizenberg, Andrew J. Black & George M. Whitesides

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

Micropatterning is a powerful method for controlling surface properties, with applications from cell biology to electronics^{1–8}. Self-assembled monolayers (SAMs) of alkanethiolates on gold and silver^{9–11}—the structures most widely used for preparing organic films with specific surface properties—are usually patterned by

partitioning the surface into regions formed from different thiols^{12–15}. Here we describe a way to pattern SAMs using a single alkanethiol on substrates consisting of regions of different topography: planar islands of one metal on the surface of a second (which may be different from or the same as the first). These topographically patterned SAMs consist of three regions: two planar surfaces and a transition region between the two. The characters of the SAMs on these three regions were inferred from images of three structures that form on them: condensation figures, patterns of crystals of CaCO₃ and regions of selective etching. The transition region is more active in the processes generating these structures than either of the two planar regions, and we propose that this activity is due to the relatively high disorder in the organic film there. We believe that this ability to

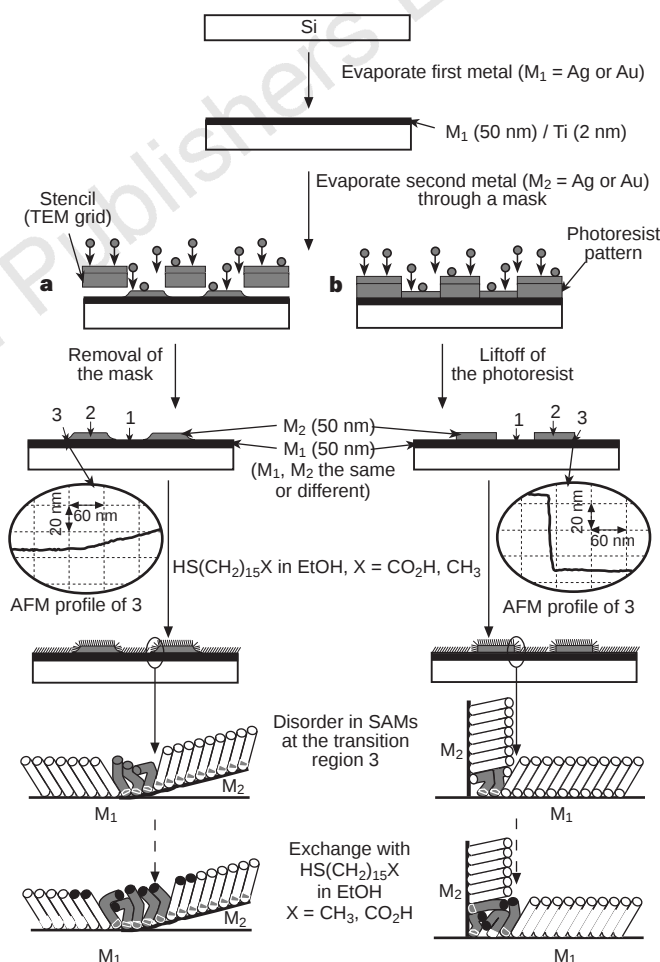


Figure 1 Schematic representation of the procedures used for fabrication of SAMs supported on topographically patterned metal surfaces. **a**, Patterned deposition of the overlayer of a metal on the surface of another metal (which might be the same or different) through a stencil. **b**, Area-selective deposition of the overlayer of a metal on the surface of another metal protected by a pattern in photoresist, followed by lift-off of the photoresist. The height profiles of the micropatterned surfaces recorded using atomic force microscopy (AFM) show that procedure **b** generates sharper edges for the patterned features than procedure **a**. The details of the surface in the transition region at the level of the grain and domain sizes and boundaries were not studied. The structure of the SAM shown is schematic. In the experiments involving formation of condensation figures, we allowed the thiol in the transition region of the SAM to exchange with a different thiol in solution. To form the patterns, silicon wafers (test grade, n or p type; Silicon Sense, Nashua, NH) were coated with 2.5 nm of Ti to promote adhesion, and then with 50 nm of metal (Ag or Au) using an electron beam evaporator and a stencil mask or patterned photoresist.

control the local disorder in a SAM with high resolution will be important in controlling processes such as nucleation, wetting, adhesion and etching on scales of below 50 nm to 5 μ m.

In this approach to micropatterning, we fabricate patterned substrates by evaporating one metal (Au or Ag) onto the surface of a second (Au or Ag) through a stencil mask or a patterned photoresist (Fig. 1). We describe these surfaces using the nomenclature M_2 -on- M_1 ; that is, M_2 (the second metal) evaporated on M_1 (the first). We formed SAMs on these patterned substrates by immersing them in a 10 mM solution of $HS(CH_2)_{15}X$ ($X = CH_3$ or CO_2H) in ethanol for 1 h. SAMs of the same alkanethiol formed on different metals differ in structure^{9–11}, and we expected, therefore, that the SAM at the border between two different metals would have a distinct and probably less-ordered structure (region 3 in Fig. 1) than the larger planar regions (regions 1 and 2 in Fig. 1). It seemed possible that the region of SAM at sharp steps in the topography of the same metal might also have abnormal structure.

To test the hypothesis that the transition region would be disordered, the topographically patterned substrates supporting one thiolate—either methyl- or acid-terminated—were allowed to soak in a solution of the second (10 mM). We expected the disordered regions of SAMs to be more labile than the ordered regions, and thus to exchange selectively^{14–16}. We examined the wettability of the surfaces using condensation figures^{17–19} to visualize modifications in the SAM resulting from this exchange (Fig. 2). The best visualization of the transition region was achieved when we used long exchange (~ 10 h) with the alternative thiol. When a methyl-terminated SAM was allowed to exchange with the acid-terminated thiol, condensation figures on the modified surface consisted of elongated drops of water at a hydrophilic border between the two metals (Fig. 2a–c). When the SAMs were formed in the reverse order—first from the acid-terminated thiol and then

exchanged with methyl-terminated thiol—the pattern of wetting was also reversed: water drops nucleated everywhere except at a hydrophobic border between the metals (Fig. 2d, e). The condensation figures outlining transition regions were particularly pronounced when M_1 was not the same as M_2 (Fig. 2b, c, e). The wetting of surfaces that were not subjected to exchange was largely homogeneous across the surface (Fig. 2a–e), with insignificant preference for nucleation at the boundaries when M_1 was not the same as M_2 presumably due to the pinning of water drops by the local disorder in the SAM. Exchange experiments using substrates prepared by deposition of M_2 at an angle of $\sim 20^\circ$ from the normal show that the size and shape of the exchanged region—which is related, we presume, to the structure of the disordered SAM—depend on the topology of the boundary between the two metals (Fig. 2c).

These preferential exchange experiments establish that there are three distinct regions in the SAMs supported on micropatterned mixed metal substrates, and suggest that one of these regions is a narrow region of a labile, disordered SAM at the border between the two metals. So this technique provides, to our knowledge, a new concept in patterning surfaces—controlling the local disorder in SAMs on a submicrometre scale by manipulating the topography and composition of the supporting metals. The boundaries between the two metals present disordered regions on the surface, and can be thus used to pattern processes that are sensitive to surface structure. For processes that require the formation of narrow features, the best system is a substrate in which M_1 is the same as M_2 , patterned using a photoresist mask to produce vertical walls in the relief. For processes that do not require the formation of narrow features, a combination of different metals gives a more pronounced effect. We demonstrate two further examples of this method: nucleation of crystal growth, and etching of the underlying metal.

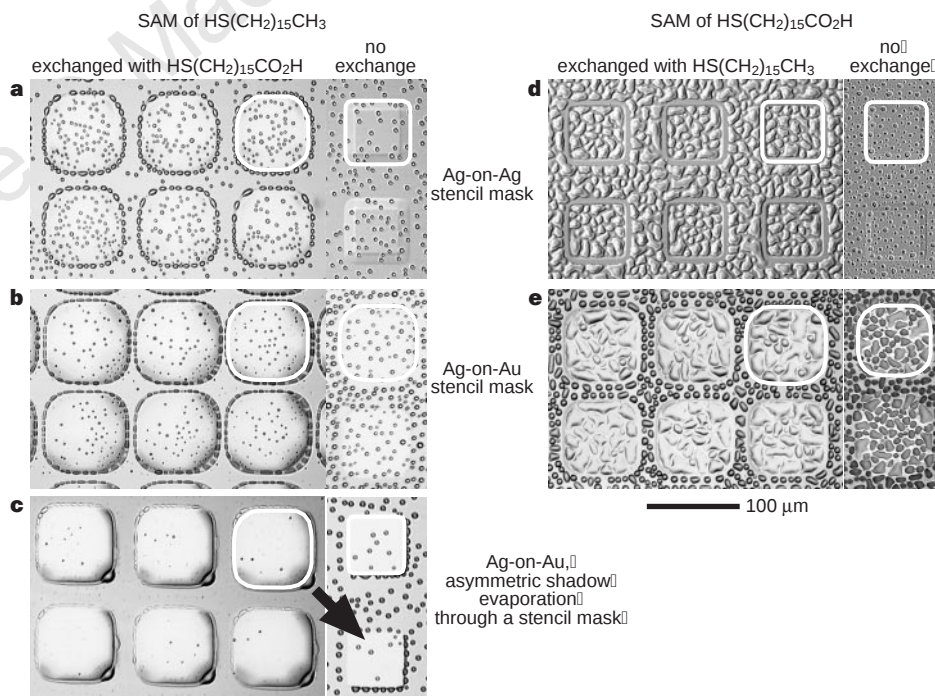


Figure 2 Optical micrographs of the condensation figures (CFs) formed on SAMs of alkanethiols supported on micropatterned metal substrates. These substrates were fabricated by depositing silver (50 nm) on silver (**a** and **d**), or silver (50 nm) on gold (**b**, **e** and **c**) using a stencil (a TEM grid). The geometries of the islands of the second metal (shown in overlays) depend on the distance between the mask and the substrate and on the angle between the beam and substrate. **a–c**, CFs formed on SAMs of $HS(CH_2)_{15}CH_3$ exchanged with $HS(CH_2)_{15}CO_2H$; **d**, **e**, CFs formed on SAMs of $HS(CH_2)_{15}CO_2H$ exchanged with $HS(CH_2)_{15}CH_3$. The rightmost columns

of each pattern show CFs formed on the control patterned substrates supporting corresponding SAMs that were not subjected to exchange. The surfaces were derivatized initially by exposing them to a solution of one thiol (10 mM in ethanol) for 30 s. The substrates were rinsed with ethanol and allowed to soak in a solution of the second thiol for 10 h. We observed CFs directly by breathing on the samples under an optical microscope. The substrate in **c** was prepared by asymmetric shadow evaporation of Ag through a stencil mask—the direction of the beam (indicated by an arrow) was $\sim 20^\circ$ from the normal to the surface.

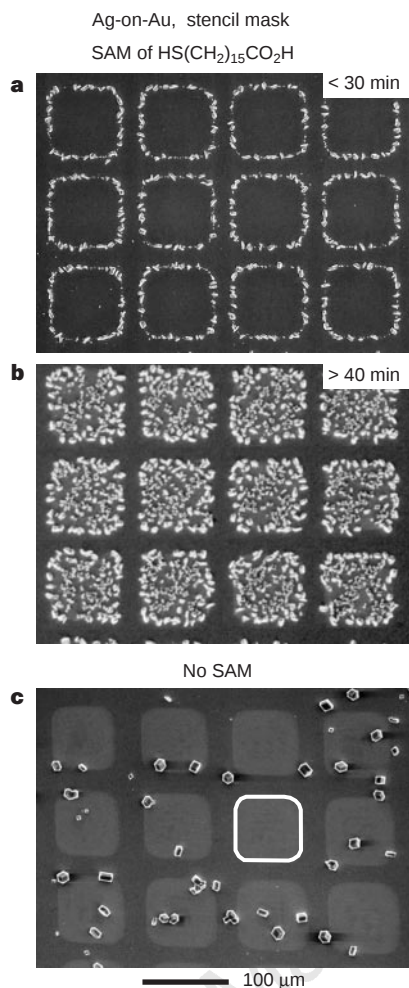


Figure 3 Scanning electron micrographs of calcite crystals grown on topographically patterned surfaces fabricated by depositing Ag (50 nm) on Au through a stencil mask followed by formation of SAMs. The overlay shows the outline of the pattern. **a**, **b**, Patterns of calcite crystals formed on SAMs of $\text{HS}(\text{CH}_2)_{15}\text{CO}_2\text{H}$ supported on micropatterned metal surfaces. For short times of crystallization (<30 min), calcitic outlines of the underlying patterns are formed (**a**). For longer times of crystallization (>40 min), preferential filling of the Ag regions with crystals takes place (**b**). **c**, Non-patterned growth of calcite induced by bare substrates supporting no SAM. The patterned metal surfaces derivatized with a SAM of $\text{HS}(\text{CH}_2)_{15}\text{CO}_2\text{H}$ were used directly in the crystallization experiments (no exchange was performed). The substrates were placed with the pattern down into a 25 mM CaCl_2 solution in a closed desiccator containing vials of solid $(\text{NH}_4)_2\text{CO}_3$.

Crystallization is highly sensitive to the structure and order of the surfaces that induce nucleation^{20–22}. We used the formation of calcium carbonate as a model system^{22–24}. The micropatterned Au-on-Ag and Ag-on-Au substrates supporting a CO_2H -terminated SAM induce patterned crystallization (Fig. 3). The crystals nucleate first at the interface between the two metals (Fig. 3a), and then fill the Ag region (Fig. 3b). The bare, micropatterned metal surfaces (with no SAM) do not induce preferential nucleation anywhere on the surface (Fig. 3c); this observation indicates that patterned nucleation arises from differences in the structure of the patterned SAM and not directly from the topography of the surface. Figure 3 suggests that these techniques may provide a simple method to pattern crystal growth²⁵. The fabrication of crystalline outlines (Fig. 3a) is especially striking, as these structures cannot be produced using other methods of crystallization.

The hypothesis that the transition regions in the SAMs are

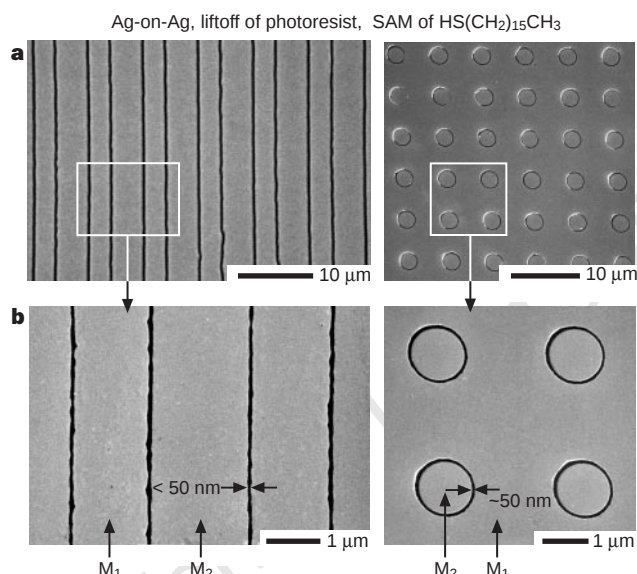


Figure 4 Selective etching of micropatterned substrates fabricated by depositing Ag (50 nm) on a Ag film through a layer of patterned photoresist, followed by lift-off. Left panels, 2-μm-wide lines; right panels, a square array of circles with 1.5 μm diameter. **a**, Low magnification SEM, showing the homogeneity of the pattern. **b**, Trenches (~50 nm wide) in the metal substrates fabricated by etching for 10 s. Substrates were exposed to a $\text{HS}(\text{CH}_2)_{15}\text{CH}_3$ solution (10 mM in ethanol) for 1 h and rinsed with ethanol, then etched with an aqueous solution of ferrocyanide (0.1 M), ferricyanide (0.001 M) and sodium thiosulphate (0.1 M).

disordered suggests that these regions expose the underlying metal to etching^{15,26}. When Ag-on-Ag or Au-on-Ag surfaces supporting a SAM of $\text{HS}(\text{CH}_2)_{15}\text{CH}_3$ were etched with an aqueous solution of potassium ferricyanide, dissolution of silver initiated preferentially at the border between the two metals (Fig. 4a). In control experiments with bare, topographically patterned substrates supporting no SAM, corrosion was uniform across the surface, and was not selective at the edges of the metal patterns. We suggest, again, that the specificity of etching at the transition regions results primarily from the disorder of the SAMs. The ability to etch the disordered regions in the SAM selectively provides a method to form small features in metal films^{1–3}. The structures generated by etching topographically patterned surfaces supporting SAMs are smaller than the original patterns used for their generation, and their feature sizes can be controlled by varying the etching time. Figure 4b shows trenches ~50 nm wide in the metal substrates fabricated by short (10-s) etching, using two sample patterns of lines (Fig. 4b, left) and circles (Fig. 4b, right).

The ability to control the local disorder in SAMs at the nanometre scale therefore provides an approach to high-resolution patterning of surfaces, and constitutes a convenient method for size reduction: features less than 50 nm wide can be generated at the edges of micrometre-sized patterns. □

Received 17 April; accepted 9 June 1998.

- Moreau, W. M. *Semiconductor Lithography: Principles and Materials* (Plenum, New York, 1988).
- Smith, H. I. A Review of Submicron Lithography. *Superlat. Microstruct.* **2**, 129–133 (1986).
- Cerrina, F. & Marrian, C. A path to nanolithography. *MRS Bull.* **12**, 56–60 (1996).
- Kim, E., Xia, Y. & Whitesides, G. M. Polymer microstructures formed by moulding in capillaries. *Nature* **376**, 581–584 (1995).
- Berggren, K. K. *et al.* Microlithography by using neutral metastable atoms and self-assembled monolayers. *Science* **269**, 1255–1257 (1995).
- Chen, C. S., Mirksich, M., Huang, S., Whitesides, G. M. & Ingber, D. E. Geometric control of cell life and death. *Science* **276**, 1425–1428 (1997).
- Bunker, B. C. *et al.* Ceramic thin-film formation on functionalized interfaces through biomimetic processing. *Science* **264**, 48–55 (1994).
- Hu, J. *et al.* Using soft lithography to fabricate GaAs/AlGaAs heterostructure field effect transistors. *Appl. Phys. Lett.* **71**, 2020–2022 (1997).
- Ulman, A. *Introduction to Thin Organic Films: From Langmuir-Blodgett to Self-Assembly* (Academic, Boston, 1991).

10. Nuzzo, R. G., Dubois, L. H. & Allara, D. L. Fundamental studies of microscopic wetting on organic surfaces. 1. Formation and structural characterization of a self-consistent series of polyfunctional organic monolayers. *J. Am. Chem. Soc.* **112**, 558–569 (1990).
11. Laibinis, P. E. *et al.* A comparison of the structures and wetting properties of self-assembled monolayers of *n*-alkanethiols on the coinage metal surfaces Cu, Ag, Au. *J. Am. Chem. Soc.* **113**, 7152–7167 (1991).
12. Dulcey, C. S. *et al.* Deep UV photochemistry of chemisorbed monolayers: patterned coplanar molecular assemblies. *Science* **252**, 551–554 (1991).
13. Tiberio, R. C. *et al.* Self-assembled monolayer electron beam resist on GaAs. *Appl. Phys. Lett.* **62**, 476–478 (1993).
14. Kumar, A., Abbott, N. A., Kim, E., Biebuyck, H. A. & Whitesides, G. M. Patterned self-assembled monolayers and meso-scale phenomena. *Acc. Chem. Res.* **28**, 219–226 (1995).
15. Xia, Y., Zhao, X. M. & Whitesides, G. M. Pattern transfer: self-assembled monolayers as ultrathin resists. *Microelectr. Eng.* **32**, 255–268 (1996).
16. Larsen, N. B., Biebuyck, H., Delamarche, E. & Michel, B. Order in microcontact printed self-assembled monolayers. *J. Am. Chem. Soc.* **119**, 3017–3026 (1997).
17. Abbott, N. L., Folkers, J. P. & Whitesides, G. M. Manipulation of the wettability of surfaces on the 0.1- to 1-micrometer scale through micromachining and molecular self-assembly. *Science* **257**, 1380–1382 (1992).
18. López, G. P., Biebuyck, H. A., Frisbie, C. D. & Whitesides, G. M. Imaging of features on surfaces by condensation figures. *Science* **260**, 647–649 (1993).
19. Kumar, A. & Whitesides, G. M. Patterned condensation figures as optical diffraction gratings. *Science* **263**, 60–62 (1994).
20. Landau, E. M., Levanon, M., Leiserowitz, L., Lahav, M. & Sagiv, J. Transfer of structural information from Langmuir monolayers to three-dimensional growing crystals. *Nature* **318**, 353–356 (1985).
21. Stupp, S. I. & Braun, P. V. Molecular manipulation of microstructures: biomaterials, ceramics, and semiconductors. *Science* **277**, 1242–1248 (1997).
22. Heywood, B. R. & Mann, S. Template-directed nucleation and growth of inorganic materials. *Adv. Mater.* **6**, 9–20 (1994).
23. Berman, A. *et al.* Total alignment of calcite at acidic polydiacetylene films: cooperativity at the organic-inorganic interface. *Science* **269**, 515–518 (1995).
24. Weiner, S. & Addadi, L. Design strategies in mineralized biological materials. *J. Mater. Chem.* **7**, 689–702 (1997).
25. Mann, S. & Ozin, G. A. Synthesis of inorganic materials with complex form. *Nature* **382**, 313–318 (1996).
26. Zhao, X.-M., Wilbur, J. L. & Whitesides, G. M. Using two-stage chemical amplification to determine the density of defects in self-assembled monolayers of alkanethiols on gold. *Langmuir* **12**, 3257–3264 (1996).

Acknowledgements. This work was supported in part by the ONR and DARPA. It used MRSEC Shared Facilities supported by the NSF.

Correspondence and requests for materials should be addressed to G.M.W. (e-mail: gwhitesides@gmwgroup.harvard.edu).

Decadal variability in the outflow from the Nordic seas to the deep Atlantic Ocean

Sheldon Bacon

Room 256/43, Southampton Oceanography Centre, Empress Dock, Southampton SO14 3ZH, UK

The global thermohaline circulation is the oceanic overturning mode, which is manifested in the North Atlantic Ocean as northward-flowing surface waters which sink in the Nordic (Greenland, Iceland and Norwegian) seas and return southwards—after overflowing the Greenland–Scotland ridge—as deep water. This process has been termed the ‘conveyor belt’, and is believed to keep Europe 5–8 °C warmer than it would be if the conveyor were to shut down¹. The variability of today’s conveyor belt is therefore an important component of climate regulation. The Nordic seas are the only Northern Hemisphere source of deep water and a previous study³ has revealed no long-term variability in the outflow of deep water from the Nordic seas to the Atlantic Ocean. Here I use flows derived from hydrographic data to show that this outflow has approximately doubled, and then returned to previous values, over the past four decades. I present evidence which suggests that this variability is forced by variability in polar air temperature, which in turn may be connected to the recently reported Arctic warming⁴.

The water which overflows the shallow (500–700 m) Greenland–Scotland ridge east and west of Iceland descends to depths around 3,000 m in the northern North Atlantic, and follows paths which are well known^{5–7}, but the prevailing view of the temporal variability of

the magnitude of the flow has been that it possesses no significant variability on timescales longer than about a fortnight³. The Nordic seas and the Arctic have been seen as a black box or reservoir whose exchanges with the rest of the world ocean are governed solely by hydraulics because shallow convective processes keep the waters at the sill depth ‘topped up’. Following the evidence presented below, it appears that shallow processes are responsive to atmospheric forcing in such a way that the Nordic seas exert important control on the variability of the conveyor belt.

Water from the Nordic seas which overflows the Greenland–Scotland ridge east of Iceland travels round the Mid-Atlantic Ridge (MAR) by way of the Charlie–Gibbs fracture zone; in the northwest of the Irminger basin, it joins the water which overflows west of Iceland through Denmark Strait. South of Denmark Strait, these waters form the Deep Western Boundary Current (DWBC), which has characteristic chemical and dynamical signatures^{5–7}. In particular, the vertical current shear near the bottom is very high, so a large part of the flow can be measured with the shear derived from hydrographic data. Without knowledge of the absolute current, one can create a 40-year history of the DWBC from archive data by using the transport referenced to a constant zero-velocity surface as an index of the strength of the DWBC.

I have analysed 22 sections, the earliest from 1955, the most recent from August 1997 (Fig. 1). They lie in the vicinity of Cape Farewell, and radiate between southwards (at ~43° W) and eastwards (at ~60° N) from Cape Farewell across the Irminger Sea. The core of the DWBC spans ~200 km of downstream distance over the sections, with the exception of the section at 62° N which was included to fill an important time gap. Being further north than the rest, the flux estimate from there may not be quite consistent with the others. The sections are far enough downstream from Denmark Strait that the DWBC is flowing nearly parallel to the isobaths and not significantly changing strength by entrainment over much of the span of the sections⁸, with the possible exception of the 62° N section. The DWBC volume flux is consistently and correctly estimated in spite of differently angled tracks of the sections across the current, assuming no significant deep flow normal to the central axis of the basin^{8,9}.

Calculating (as previous authors have done) the DWBC as transport below the density surface $\sigma_0 = 27.8$, with zero velocity at 1,000 m, a picture of DWBC magnitude and variability emerges (Fig. 2). Low transports (nine estimates) appear in the 1950s and 1960s, high (four estimates) in the 1970s and 1980s, and low again (nine estimates) in the 1990s. Is this noise or a signal? First, there is no bias with section orientation or season. Second, were it true for the present data set as for previous ones that the dominant variability period is a few days, this set would be the product of a random process. This hypothesis may be tested by dividing the data into three groups as above and comparing the first with the second and the second with the third (using a non-parametric ranking or mean slippage test such as Wilcoxon’s¹⁰). In both cases, the chance that the pattern is random is <0.5%.

The relative invariance of the referenced transport signal on short time and space scales is tested by examining two groups of sections: four sections taken over four weeks in spring 1966, and five sections taken over five months around summer 1991 (Fig. 1, left and right insets). These two groups have mean transports (standard errors in parentheses) of 5.8 (0.2) Sv (first group; 1966), and 4.5 (0.2) Sv (second group; 1991) (1 Sv = 10⁶ m³ s⁻¹). The ‘short-scale’ range (0.2 Sv) is an order of magnitude less than the decadal range. Therefore I treat the pattern as a signal.

What of the transport due to the reference level current? First, recent modelling studies^{11,12} indicate that the Irminger basin circulation is not strongly dependent on variations in the wind stress. The referenced transport signal is unlikely to have been altered by the wind-driven circulation in such a way as to annul exactly the observed variability. Second, absolute transports are available in two

cases. The first, from the 1978 RV *Hudson* section, was generated^{3,6,9} by merging the velocity shear with the mean of six weeks of current-meter data to produce the presently accepted (high) DWBC value of ~ 13 Sv. Six weeks is long enough to produce an adequate mean, given maximum variability around periods of a few days. The second, from the 1991 RV *Charles Darwin* sections, was calculated by an inverse method applied to the sections' hydrographic data¹³; a low transport of 6 Sv resulted, but it cannot be rejected as a consequence of the 'minimum energy' inversion constraint because the solution is forced in the vicinity of the DWBC by shipboard acoustic Doppler current profiler data, and subjected to sensitivity testing. Therefore, I argue that both estimates are robust.

The nine 1990s referenced transport estimates have a mean (standard error) of 4.3 (0.2) Sv, so the absolute inversion estimate, based on a referenced transport of 4.1 Sv, is not unusual on that account. Furthermore, the resulting strength of the sub-polar gyre (27 Sv) was similar to accepted values (~ 30 Sv)^{14–16}. The layer of water from $\sigma_\theta = 27.8$ to the bottom occupies about one-third of the water column, so to 'inflate' the DWBC by increasing the transport due to the current at the reference level (to raise the absolute transport from 6 to 13 Sv) one should put an extra 7 Sv in the DWBC, which would require 21 Sv extra in the whole water column. In consequence, the south-going Atlantic transports by Greenland and the north-going North Atlantic Current by Europe would be in excess of 50 Sv, approximately double the accepted values. This

would upset the calculation of net fluxes of heat, fresh water, and so on. It is unlikely that the decline from the 1978 to the 1991 absolute transport estimates can be explained by erroneous reference level currents in the latter.

In both absolute transport cases, I split the DWBC into the referenced transport (due to the intense, highly sheared deep flows) and transport due to the actual current at the reference level. For 1978, $13 \text{ Sv} = 8 + 5 \text{ Sv}$, and for 1991, $6 \text{ Sv} = 4 + 2 \text{ Sv}$ (respectively), so over half (4 Sv) of the 'lost' absolute transport of 7 Sv is to be found in the changed referenced transport. The reduction of 3 Sv due to the current at the reference level probably results from reduction in entrainment between Denmark Strait and Cape Farewell, caused by reduced outflow from the Nordic seas. This would be consistent with the implications of model sensitivity studies¹².

I now try to reconcile the observed changes in the DWBC with the prevailing model^{3,6} of the downstream development of Nordic seas outflow, which is based on current-meter data¹⁷ from 1973 south of the Denmark Strait sill (3 Sv below $\sigma_\theta = 27.8$), arrays^{3,6} from 1986 to 1991 between ~ 65 to 63°N (5 Sv increasing downstream to 11 Sv), and an array⁹ from 1978 south of Cape Farewell (13 Sv). These agree reasonably with the best model of the Nordic seas outflows at

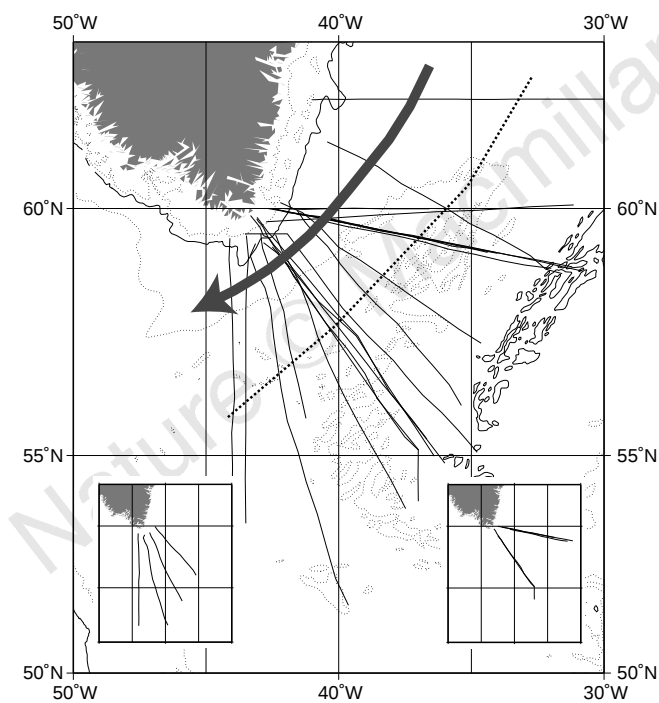


Figure 1 The 22 sections used in the analysis. Bathymetry is illustrated with the 200-m, 1500-m (solid) and 3000-m contours. The heavy dashed line defines the middle of the basin (see Fig. 2 legend). The wide grey arrow shows the approximate location of the core of the DWBC. The left inset shows the subset of four sections from 1966; the right inset shows the subset of five sections from 1991. The research cruises from which the sections are derived are specified below by *Ship* (month/year); in two cases, more than one section in a given month has been used from the relevant expedition, indicated by x. The sections are: *Anton Dohrn* (6/55, 4/58, 9/58), *Erika Dan* (3/62); *Hudson* (3/66 $\times$ 3, 4/66, 2/67, 4/78), *Knorr* (7/81, 7/83), *Poseidon* (5/88), *Tyro* (4/91), *Hudson* (5/91), *Charles Darwin* (8/91 $\times$ 2), *Meteor* (9/91), *Valdivia* (9/92), *Meteor* (11/94), *Discovery* (10/96, 8/97). The 1958 cruises are part of the International Geophysical Year (IGY) survey; the 1981 data are part of the Transient Tracers in the Ocean (TTO) program; all 1990s cruises are part of the World Ocean Circulation Experiment (WOCE).

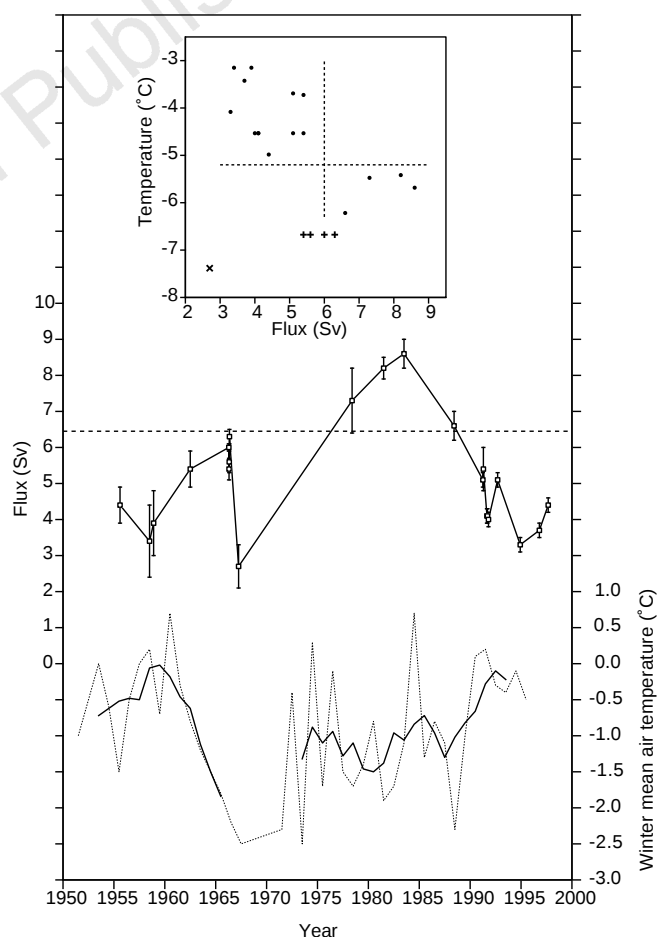


Figure 2 The decadal time series. The time series of DWBC volume fluxes (Sv; upper series), including a transport value (dashed line) which separates the series into two groups. Details of the calculations are given in Methods. The lower series are the Jan Mayen winter (December to March) mean air temperatures (dashed line), overlaid by a three-point running mean (heavy line). Note the extreme cold in the late 1960s. The inset shows overflow flux plotted against the mean air temperature of the three preceding winters (that is, the running mean lagged by one year). The dots show the bulk of the points, which are divided by the dashed lines into cold/high outflow and warm/low outflow groups. The 1966 (+) and 1967 (x) points fall outside this grouping. Note that the two *Charles Darwin* values in 1991 are of the same transport, and separated in time by a week, so that the points overlie.

present available⁸. There is no difficulty in accommodating a strong DWBC in 1973 in Fig. 2, but it is necessary to consider in more detail events in early 1991.

Only one of the three later arrays was deployed in 1990–91. Plots of monthly mean near-bottom current near the current core for two instruments are shown in ref. 3, Fig. 11. One (9008–3) shows fairly steady currents, the record ending in April 1991. The other (9010–3) ends in June 1991, and the last three monthly means are all decreasing from the March value. The hydrographic data for 1991 begin in April; so I conclude that the 1986–91 measurement effort just caught the beginning of the decrease in the DWBC, but this (in context) small piece of evidence was not significant in the light of five preceding near-steady years.

To identify the source of the DWBC decadal transport variability, I first examine the effects of local changes in the structure of the northern North Atlantic. Large secular changes have been documented^{18–20} in the overlying water mass, Labrador Sea Water (LSW). Water leaving the Nordic seas passes over the Greenland–Scotland sill; the plume which descends into the deep northern North Atlantic basins passes through the LSW layer. None of the smaller pre-1990 density excursions of the LSW coincide with DWBC changes, but the largest excursion, of strongly increasing density, does coincide with the post-1990 DWBC decrease. How might a change in entrainment of LSW into the descending overflow plume near the sill affect the DWBC strength? Entrainment is inversely proportional to the Richardson number²¹, and so is inversely proportional to the density contrast between the plume and the surrounding water. An increase in LSW density would decrease the density contrast, increasing entrainment, and so increasing the DWBC, but the reverse is observed, so this seems an unlikely cause.

So the DWBC changes are most probably a result of changed output from the Nordic seas. Is it possible to find a cause? The water which flows over the sill is a Nordic seas intermediate water mass (Arctic Intermediate Water, AIW). Several models describe its formation: by winter heat loss to the atmosphere in the centre of the Iceland and Greenland Sea gyres²²; in the east of the Nordic seas²³ over the Norwegian Coastal Current; or by isopycnal mixing²⁴ in the west, in the eastern boundary of the East Greenland Current (where flow from the Arctic proper, via Fram Strait, is implicated). I examine these (they may all be implicated) by inspecting the winter mean air temperature over the Nordic seas, represented by measurements from Jan Mayen Island (71°N, 8.5°W), ideally placed at the junction of the ridges separating the Iceland, Greenland and Norwegian basins. Figure 2 shows the air-temperature history and (inset) the overflow flux plotted against air temperature. Most of the latter points divide into two groups: low air temperature associated with high outflow, and vice versa. This suggests that the air temperature sets AIW density and/or layer thickness, which in turn cause the observed changes in outflow flux, and in the sense predicted by hydraulic exchange theory. A link between atmospheric temperature and actual AIW temperature is made in ref. 25, plate 1 (central Greenland Sea temperature, 1952–95, surface to bottom), where the pattern of Fig. 2 is repeated: in the 1950s and 1960s, the AIW (200–600 m depth) is warm (low outflow), in the 1970s and 1980s it is cold (high outflow), and in the 1990s, it is strongly warming (low outflow again).

Some points in Fig. 2 inset fall outside this relationship. They are of very low air temperatures and of moderate-to-low outflow flux, and originate in 1966/7. It may be possible that here a different set of rules apply to those described above. The unusually low air temperatures of the mid-1960s are associated with very high ice cover²⁶—this event is cited²⁷ as the root cause of the ‘great salinity anomaly’²⁸—so it appears that the high ice cover insulated the AIW formation region from the extreme cold of the air, resulting in low air temperature/low outflow conditions.

The phase relationships between many linked ocean–atmosphere buoyancy-driven phenomena in the North Atlantic are reviewed in

ref. 19: I believe that the results reported here are another such phenomenon. It appears that changes in the Arctic and North Atlantic atmospheric and oceanic circulation are linked, and are at present weakening the Atlantic (and so the global) thermohaline circulation. This, ultimately, could alter the way that heat is redistributed around the globe by both ocean and atmosphere. For the Greenland Sea, the accumulated heat could at present be removed by just two years of intense winter cooling²⁵. But how far does Arctic warming need to progress before the oceanic circulation changes catastrophically in the manner suggested by recent coupled ocean–atmosphere modelling efforts^{29,30}, which represent climate transitions between ‘normal’ (present) and ice-age states? We cannot yet tell whether, or how much of, the variability reported here is normal or unusual. □

Methods

The volume fluxes were calculated by imposing a level of zero motion at 1,000 m and the total flux below the density level $\sigma_0 = 27.8$ was computed. Vertical velocity shear is derived from the horizontal pressure gradient through the geostrophic relationship, so one must estimate the flux in the ‘bottom triangles’: that is, for each pair of stations, the area between the bottom of the shallower station, and the bottom of the deeper station. Projection of the deepest common velocity provides an underestimate, projection of the deepest common shear provides an overestimate. These are the bases of the flux error bars; horizontal station resolution thus governs the error estimate, so the earlier sections, generally composed of wider spaced stations than the later sections, have larger associated errors. Allowance was made in two cases for the section terminating part-way through the DWBC. No significant difference resulted from changing the zero velocity level by a few hundred metres in depth. The calculation was extended from the Cape Farewell end of each section to a line defining the middle of the Irminger basin (see Fig. 1), except where that resulted in the line ending in the middle of an eddy with a baroclinic signature, in which case the calculation was concluded to one side of the eddy. The eddies were symmetrical, so it was important merely to avoid extending the calculation into the north-going flow on the west flank of the Mid-Atlantic Ridge. The eastern edge of the DWBC in mid-basin carries much less water than the core, which lies between 2,000 and 3,000 m depth on the Greenland continental slope.

Received 3 March 1997; accepted 27 May 1998.

1. Broecker, W. S. The great ocean conveyor. *Oceanography* **4**, 79–89 (1991).
2. Warren, B. A. In *Evolution of Physical Oceanography* (eds Warren, B. A. & Wunsch, C.) 6–41 (MIT Press, Cambridge, MA, 1981).
3. Dickson, R. R. & Brown, J. The production of North Atlantic Deep Water: sources, rates and pathways. *J. Geophys. Res.* **99**, 12319–12341 (1994).
4. Macdonald, R. Awakenings in the Arctic. *Nature* **380**, 286–287 (1996).
5. Swift, J. H. The circulation of the Denmark Strait and Iceland–Scotland overflow waters in the North Atlantic. *Deep-Sea Res.* **31**, 1339–1355 (1984).
6. Dickson, R. R., Gmitrowicz, E. M. & Watson, A. J. Deep-water renewal in the northern North Atlantic. *Nature* **344**, 848–850 (1990).
7. McCartney, M. S. Recirculating components to the deep boundary current of the northern North Atlantic. *Proc. Oceanogr.* **29**, 283–383 (1992).
8. Price, J. F. & Baringer, M. O. Outflows and deep water production by marginal seas. *Prog. Oceanogr.* **33**, 161–200 (1994).
9. Clarke, R. A. Transport through the Cape Farewell–Flemish Cap section. *Rapp. P.-v. Réun. Cons. Int. Explor. Mer.* **185**, 120–130 (1984).
10. Pearson, E. S. & Hartley, H. O. *Biometrika Tables for Statisticians* Vol. 2 (Biometrika Trust, London, 1976).
11. Bryan, F. O., Böning, C. W. & Holland, W. R. On the midlatitude circulation in a high-resolution model of the North Atlantic. *J. Phys. Oceanogr.* **25**, 289–305 (1995).
12. Redler, R. & Böning, C. W. Effect of the overflows on the circulation in the subpolar North Atlantic: a regional model study. *J. Geophys. Res.* **102**, 18529–18552 (1997).
13. Bacon, S. Circulation and fluxes in the North Atlantic between Greenland and Ireland. *J. Phys. Oceanogr.* **27**, 1420–1435 (1997).
14. Schmitz, W. J. & McCartney, M. S. On the North Atlantic circulation. *Rev. Geophys.* **31**, 29–49 (1993).
15. Reid, J. L. On the total geostrophic circulation of the North Atlantic Ocean: flow patterns, tracers, and transports. *Prog. Oceanogr.* **33**, 1–92 (1994).
16. Talley, L. D. North Atlantic circulation and variability. *Physica D* **98**, 625–646 (1996).
17. Ross, C. K. Temperature–salinity characteristics of the ‘overflow’ water in Denmark Strait during ‘OVERFLOW ’73’. *Rapp. P.-v. Réun. Cons. Int. Explor. Mer.* **185**, 111–119 (1984).
18. Read, J. F. & Gould, W. J. Cooling and freshening of the subpolar North Atlantic Ocean since the 1960s. *Nature* **360**, 55–57 (1992).
19. Dickson, R. R., Lazier, J. R. N., Meincke, J., Rhines, P. & Swift, J. Long-term coordinated changes in the convective activity of the North Atlantic. *Proc. Oceanogr.* **38**, 241–295 (1997).
20. Sy, A. et al. Surprisingly rapid spreading of newly formed intermediate waters across the North Atlantic Ocean. *Nature* **386**, 675–679 (1997).
21. Turner, J. S. Turbulent entrainment: the development of the entrainment assumption, and its application to geophysical flows. *J. Fluid Mech.* **173**, 431–471 (1986).

22. Swift, J. H., Aagaard, K. & Malmberg, S.-A. The contribution of the Denmark Strait overflow to the deep North Atlantic. *Deep-Sea Res.* **27**, 29–42 (1980).
23. Mauritzen, C. Production of dense overflow waters feeding the North Atlantic across the Greenland–Scotland ridge. Part 1: Evidence for a revised circulation scheme. *Deep-Sea Res.* **143**, 769–806 (1996).
24. Strass, V. H., Fahrback, E., Schauer, U. & Sellman, L. Formation of Denmark Strait Overflow water by mixing in the East Greenland Current. *J. Geophys. Res.* **98**, 6907–6919 (1993).
25. Bönisch, G., Blindheim, J., Bullister, J. L., Schlosser, P. & Wallace, D. W. R. Long-term trends of temperature, salinity, density and transient tracers in the central Greenland Sea. *J. Geophys. Res.* **102**, 18553–18571 (1997).
26. Mysak, L. A., Manak, D. K. & Marsden, R. F. Sea-ice anomalies observed in the Greenland and Labrador Seas during 1901–1984 and their relation to an interdecadal Arctic climate cycle. *Clim. Dyn.* **5**, 111–133 (1990).
27. Häkkinen, S. An Arctic source for the Great Salinity Anomaly: a simulation of the Arctic Ice–Ocean system for 1955–75. *J. Geophys. Res.* **98**, 16397–16410 (1993).
28. Dickson, R. R., Meincke, J., Malmberg, S.-A. & Lee, A. J. The “Great Salinity Anomaly” in the northern North Atlantic 1968–1982. *Prog. Oceanogr.* **20**, 103–151 (1988).
29. Rahmstorf, S. Rapid climate transitions in a coupled ocean–atmosphere model. *Nature* **372**, 82–85 (1994).
30. Webb, R. S., Rind, D. H., Lehman, S. J., Healy, R. J. & Sigman, D. Influence of ocean heat transport on the climate of the last glacial maximum. *Nature* **385**, 695–699 (1997).

Acknowledgements. I thank R. Hendry (BIO) for the Hudson 1991 data; A. Sy (BSH) for the Meteor 1994 data; R. Pollard and P. Holliday (SOC) and H. Leach (LUDO) for the Discovery 1996 data; S. Cunningham for the Discovery 1997 data; the German Oceanographic Data Centre (BSH, Hamburg) for the Poseidon 1988 data. The Jan Mayen data were obtained from the Climatic Research Unit at UEA. (All other pre-1990 data were obtained from the US NODC CD-ROM 02. All other post-1990 data were obtained from the WOCE Data Assembly Centre (BSH, Hamburg).)

Correspondence and requests for materials should be addressed to the author (e-mail: S.Bacon@soc.soton.ac.uk).

Melt to mush variations in crustal magma properties along the ridge crest at the southern East Pacific Rise

S. C. Singh*, G. M. Kent†, J. S. Collier*, A. J. Harding† & J. A. Orcutt†

* BIRPS, Bullard Laboratories, University of Cambridge, Cambridge CB3 0EZ, UK

† Ida & Cecil Green Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, California 92093, USA

The determination of along-axis variations in melt properties within the crustal axial magma chamber beneath fast spreading axes is important for understanding melt delivery from the mantle, eruption history along the ridge crest, and the process of crustal accretion. Seismic reflection images^{1–4} have shown the molten sill to be continuous along the ridge crest for many tens of kilometres with varying widths (250–4,500 m), but variations in its seismic properties and thickness have remained elusive, despite several attempts to constrain these properties^{5–7}. Here we report that the melt sill along the southern East Pacific Rise, which is about 50 m thick, undergoes abrupt changes in its internal properties, ranging from pure melt to mush. The 60-km-long ridge-crest segment near 14° 00' S is characterized by three 2–4 km sections containing pure melt embedded within a magma chamber rich in mush. These small pure melt pockets may represent a fresh supply of magma from the mantle, capable of erupting and forming the upper crust. Conversely, the 80–90% of the magma chamber which is mushy is unlikely to erupt and may influence the lower-crustal accretion.

Here we use single-ship (conventional) and two-ship multi-channel seismic (MCS) reflection data collected during the TERA experiment². The linear ridge segment near 14° S was the site of a detailed MCS investigation⁸, including a conventional seismic reflection experiment with 13 profiles shot across the rise axis, interconnected by an along-axis profile (SEPR45) running the length of the survey area. The conventional seismic profile (Fig. 1) along the ridge crest (SEPR45) shows a continuous reflection from the axial magma chamber (AMC) along the whole profile. In addition, a wide-aperture profile (WAP58) was collected along-axis to complement the single-ship profiles⁹. The maximum

source–receiver offset obtained was ~7.4 km, which provided coverage of both refracted and reflected arrivals originating at the AMC and above. The extended aperture and variety of seismic data collected along the 14° S segment provides a unique opportunity to study the molten state of the AMC.

Seismic (P- and S-wave) velocities of molten material depend on the seismic velocities of its constituent fluid and crystals, and on the way the crystals are organized¹⁰. Some laboratory measurements have been made on seismic velocities by melting basalts at atmospheric pressure^{11,12}. Although we cannot apply these results directly, they, along with theoretical studies¹⁰, provide some insight into the velocity variations as a function of the fluid content. As the fluid content increases from 0% to 100%, the P-wave velocity decreases from that of hot rock to the velocity of fluid (pure melt). The S-wave velocity also decreases with an increase in melt fraction, but it suddenly drops to zero when the percentage of crystals in the melt is small enough that the crystal aggregates are completely disconnected and are unable to support shear stress. The exact temperature and crystal fraction at which the S-wave velocity drops to zero are not well constrained. However, here we describe a melt, at least seismically, by a near-zero S-wave velocity (<0.5 km s⁻¹), and a mush by a non-zero (>1.5 km s⁻¹).

A preliminary inspection of common mid-point (CMP) gathers from the along-axis WAP58 showed the character of the AMC reflection to vary beneath different sections of the ridge (Fig. 2). At CMP 1625, the AMC is a strong event out to at least 3.2 km range whilst the AMC amplitude at CMP 2488 is weak beyond ~2 km range. The amplitude of the theoretical reflection coefficient patterns (Fig. 2) suggest the presence of melt at CMP 2488 and mush at CMP 1625. In addition to the P-wave AMC reflection, we recognise a new phase, a converted shear wave arrival within gather 2488 that we will call P_{melt}S. The detection of the converted phase is important as it rules out the possibility that the reduction of the P-wave AMC amplitude at CMP 2488 results from navigational errors. P_{melt}S is not strong in CMP 1625, as predicted from our theoretical calculations for a mushy interface, which show the converted phase to have a low amplitude.

To quantify the internal properties of the melt sill, we apply a full waveform inversion to these two CMP gathers (see Methods). The structure of the AMC is well constrained by waveform inversion (Fig. 3). The inversion results show that the AMC reflection is produced by a thin (50-m) low-velocity layer in which the P-wave velocity drops from 6.0 to 3.4 km s⁻¹ in gather 2488, and to 4.0 km s⁻¹ in gather 1625. At CMP 2488, we determine the S-wave velocity to be close to zero, while beneath CMP 1625 a non-zero (2.3 km s⁻¹) S-wave velocity in the sill is observed.

From the analysis described above it is clear that the crucial window for discriminating between the two gathers is at ~2–3 km range: if pure melt were present, the amplitude of the AMC reflection would diminish and undergo a phase change, and the converted phase would be strong. We performed a 2–3 km offset stack of the 60-km along-axis line SEPR45 (Fig. 4c, d) (see Methods). The stack indicated that the bulk of this section was underlain by a magma chamber containing mush, but at three locations, separated by 15–20 km, the sill appears to be melt-rich. We then performed partial P-wave and P_{melt}S stacks for both lines SEPR45 (Fig. 4a) and WAP58 (Fig. 4b) that were shot at different times along the same track. As expected, the stacked P_{melt}S event is strong near gather 2488, but absent near gather 1625. Contrary to this, the partial P-wave stack shows strong P-wave energy in the mush zone and weak or no energy in the pure melt zone. Taken together, these two images provide a powerful tool to map the molten state of a magma chamber. As the amplitude of the whole WAP58 stack between CMP 2750 and 3000 is weak, the lateral extent of the pure melt zone is difficult to determine from this profile alone (Fig. 4b). However, SEPR45 (Fig. 4a) suggests that the lateral extent of the pure melt zone is between 2 and 4 km. We also performed

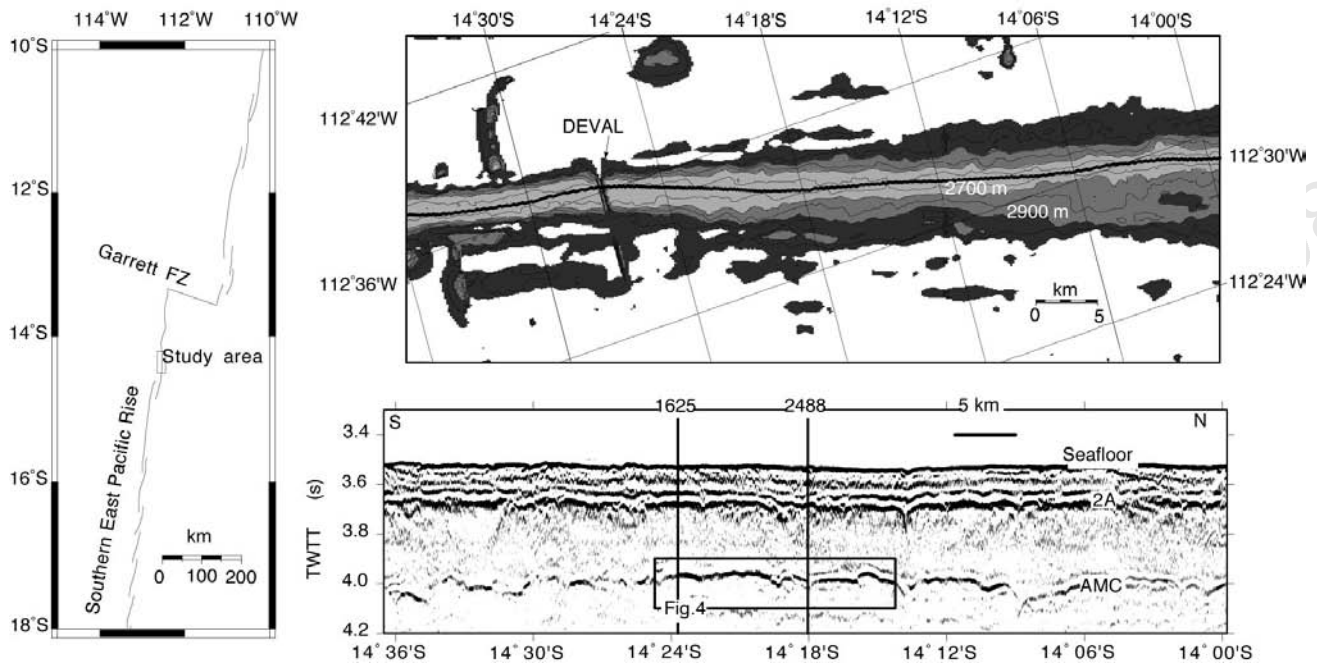


Figure 1 Study area. Left, location map of the study area. Top right, bathymetric map of the southern East Pacific Rise showing location of along-axis profile SEPR45. Bottom right, Conventionally stacked seismic line SEPR 45°. The axial magma chamber (AMC) is at ~4 s two-way travel time (TWTT). The locations of CMPs 1625 and 2488 used in the study are also marked.

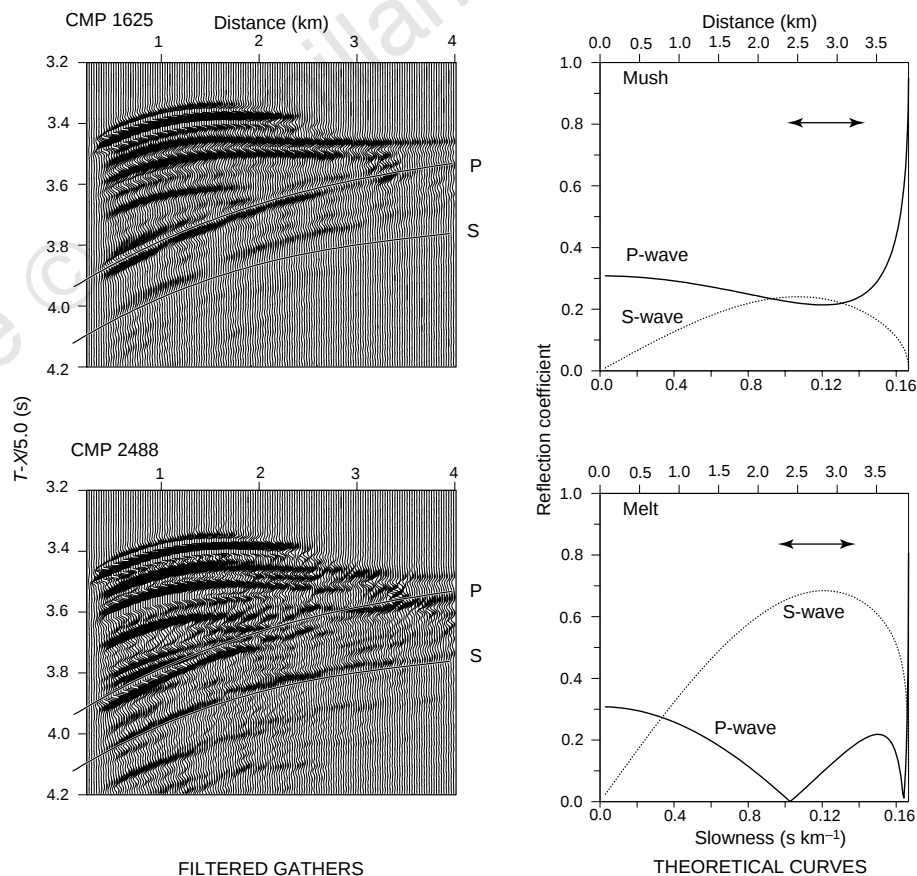


Figure 2 Common mid-point gathers and theoretical reflection coefficients. Left panels, frequency-wavenumber filtered common mid-point (CMP) gathers from WAP58 (top, CMP 1625; bottom, CMP 2488; see Fig. 1 for location). Travel-time/distance curves for the AMC (P) and P_{mel} (S) were computed from the P-wave model and Poisson ratio structure from Christeson *et al.*²⁴ Right panels, theoretically computed P-wave (solid) and S-wave (dotted) reflection coefficients (displacement) for a P-wave incident at a half-space interface. The upper layer is

solid (P-wave velocity $v_P = 6.0 \text{ km s}^{-1}$, S-wave velocity $v_S = 3.2 \text{ km s}^{-1}$, density = $2,700 \text{ kg m}^{-3}$) underlain by mush (top panel; $v_P = 3.3 \text{ km s}^{-1}$, $v_S = 2.0 \text{ km s}^{-1}$, density = $2,600 \text{ kg m}^{-3}$) and pure melt (bottom panel; $v_P = 3.3 \text{ km s}^{-1}$, $v_S = 0.0 \text{ km s}^{-1}$, density = $2,600 \text{ kg m}^{-3}$). We note that a phase reversal occurs beyond the amplitude minimum at ~2.4 km offset in the case of melt. The arrow indicates the distance window used for stacking in Fig. 4.

partial stacks of the across-axis lines in the area, all of which showed the presence of mush. The partial P-wave and S-wave stacks for the most northerly of the three melt locations identified in Fig. 4c were similar to those for the central area, and show that the pure melt zone is ~ 3 km long. (see Fig. A1 in Supplementary Information) The $P_{\text{melt}}S$ stack for the most southerly melt area did not show strong energy.

The P-wave and S-wave velocities obtained from the waveform inversion can be used to constrain the physical state of material within the AMC. The P-wave velocities determined at both locations are slightly higher than expected for pure melt, which could be interpreted as indicating the presence of some crystals in the melt sill. However, the near-zero S-wave velocity at gather 2488 indicates that even if crystals were present they are not connected; thus we suggest that the magma at gather 2488 consists of nearly pure melt. On the other hand, the P-wave velocity at gather 1625 is slightly higher (4.0 km s^{-1}) than observed at gather 2488, but the S-wave velocity is non-zero (2.3 km s^{-1}) suggesting that the crystals are interconnected as they support shear stress. We have defined material with this property within the sill as mush. From theoretical studies¹³, one can suggest that 90–95% melt could be present in the pure melt region and 40–60% in the mush zone. On the basis of studies of lava lakes, Marsh¹⁴ argues that a basalt melt with a crystallinity of $\sim 25\%$ behaves rheologically as a solid and it is unlikely to be erupted. This suggestion is consistent with the observation that samples collected at the northern East Pacific Rise tend to contain ~ 7 –10% phenocrysts¹⁵, which implies that there is less likelihood of the mush erupting, whereas the melt lens is capable of erupting.

Over the 60-km seismic section, there seem to be three zones of pure melt, each with 2–4 km of lateral extent (Fig. 4e). The pure melt may correspond to zones of fresh supply of magma from the mantle, and the mush may have gone through a cooling and

crystallization process and hence be more evolved. These lateral variations in magma-chamber structure may inhibit large-scale mixing or flow of magma along the ridge axis. Our analysis also suggests that the AMC could be segmented on a 15–20 km scale, or at least melt may be replenished from the uppermost mantle on a similar length scale (Fig. 4e). Although segmentation in the AMC has been postulated from seismological^{16,17}, morphological¹⁸ and petrological^{19,20} studies, the small-scale melt–mush segmentation is a new result, and provides an insight on the melt delivery, eruption history and crustal accretion along a spreading axis. The presence of 80–90% uneruptable mush in the magma chamber suggests that a large part of the AMC plays a passive role in the crustal accretion, and that the magmatic events such as dyking and eruptions are associated with replenishment of the AMC by magma from a deeper reservoir. This means that the pure melt pockets may be responsible for the upper-crustal formation whereas the mushy zones may influence the accretion of the lower crust. A 2–4-km long, 1,000-m-wide and 50-m-thick pure melt region could cool and crystallize to form mush in a few decades²¹, and may explain the temporal variations observed from petrological studies²⁰. On a short timescale (a decade or two), the melt supply from the mantle could be episodic²², filling only the pure melt regions at a scale-length of tens of kilometres along axis. However, on a longer timescale (100–1,000 years), it should be considered a steady-state process where pure melt segments shift along axis to enable a more steady-state style of accretionary emplacement.

Central to obtaining these new results was the determination of the detailed seismic structure of the melt sill at two neighbouring locations, using waveform inversion applied to wide-aperture data. Of equal importance is the identification of a previously unrecognized shear phase that originates at the melt interface, providing additional constraints on the determination of melt sill properties. Quantification of the physical properties could be obtained using

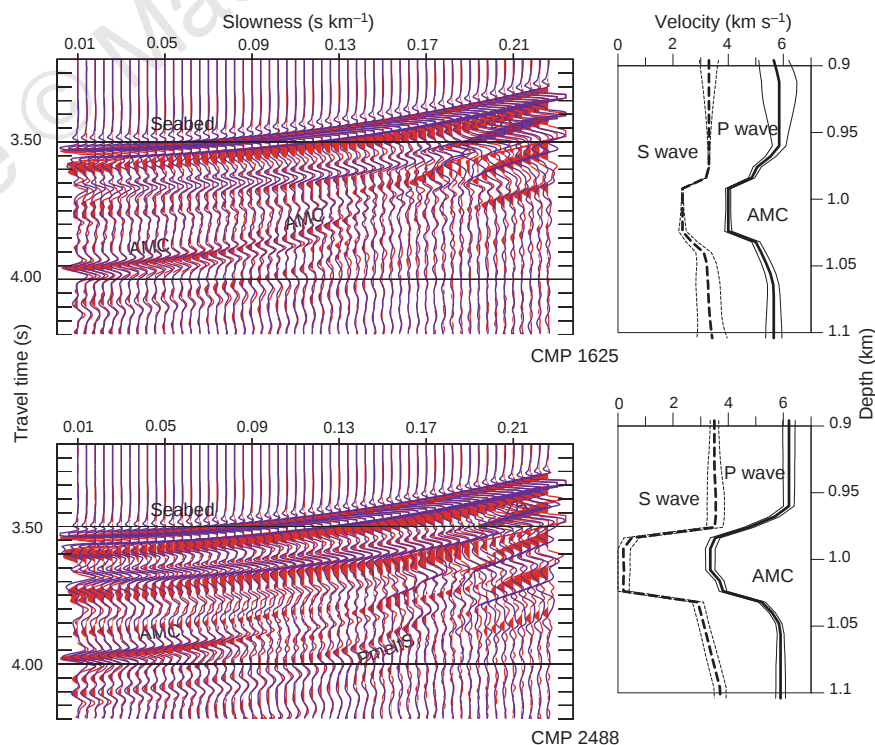


Figure 3 Results of waveform inversion. Left panels, observed (red) and synthetic (blue) data in the tau-slowness domain for WAP58, CMP 1625 (top) and CMP 2488 (bottom). We note the profound difference between the seismic signature of the AMC and the presence of the $P_{\text{melt}}S$ phase in CMP 2488. The waveform fit is extremely good within each CMP gather. (For data residuals, see Figs A3 and A4 in Supplementary Information.) Right panels, P-wave and S-wave velocities after

the inversion at CMP 1625 (top) and CMP 2488 (bottom). Thick solid (P-wave) and dashed (S-wave) lines are our best velocity estimates, and thin solid and dashed lines are one standard deviation error in the final velocity models. These errors were estimated from the Hessian matrix²⁵. The melt sill is ~ 50 m thick at the two locations. The S-wave velocity at CMP 2488 is near zero, and is non-zero at CMP 1625.

the waveform inversion technique, which can be thought of as analogous to a borehole sonic log, and in conjunction with $P_{\text{melt}}S$ and partial P-wave images it could be used to extend the pinpoint measurements in more continuous fashion. If the properties of the melt in the magma chamber changes with time, our method could be used to determine the temporal variation by analysing data

collected over a decadal timescale. More importantly, this technique could be used to predict which melt-rich sections of the ridge crest are likely to erupt next. □

Methods

Waveform inversion. The seismic structure above the magma-chamber

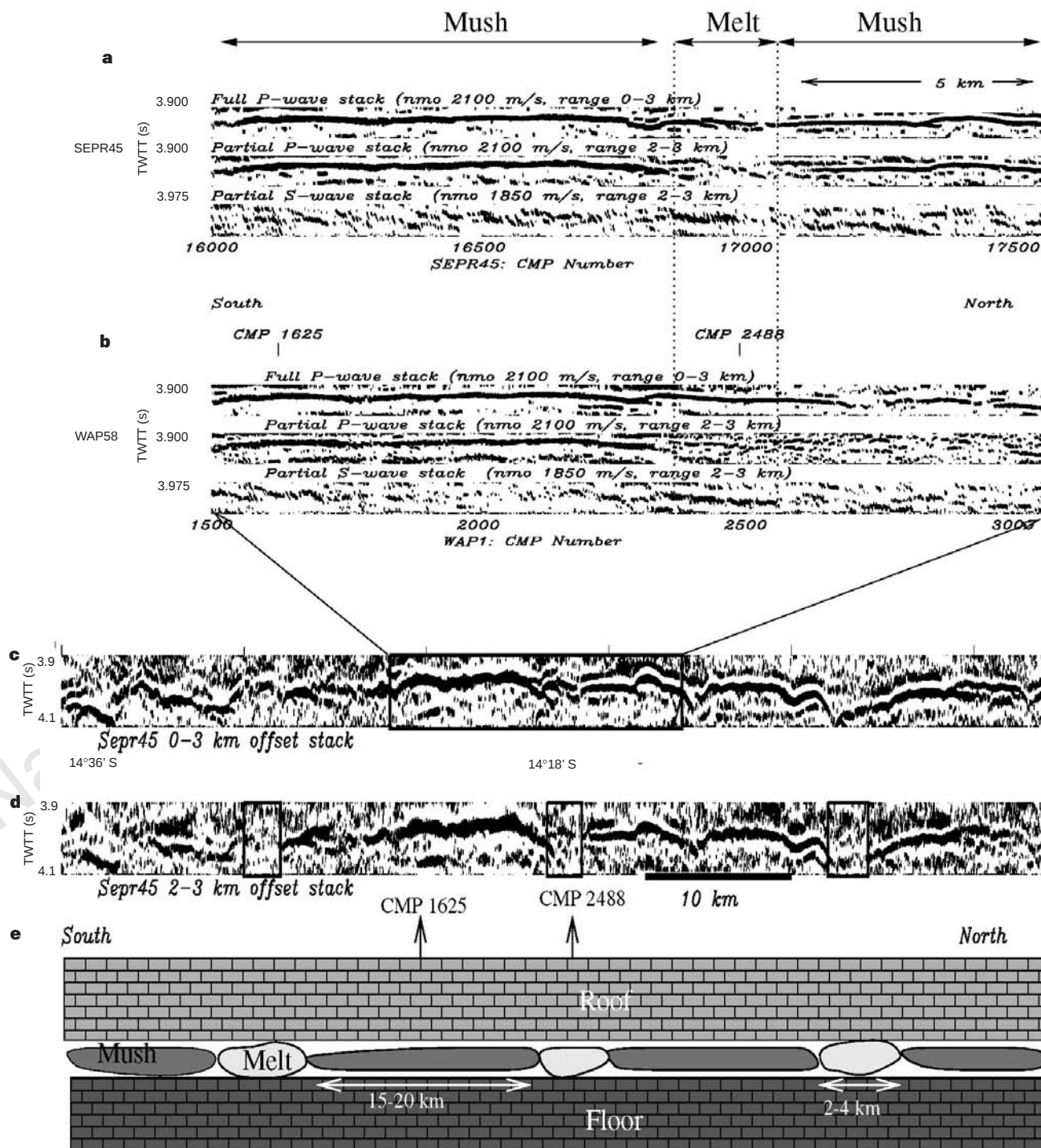


Figure 4 Results of stacking of final interpretation. Full P-wave, partial P-wave and S-wave stack for both SEPR45 (a) and WAP58 (b), conventional stacked seismic image of SEPR45 of traces in the range 0–3 km offset (c). The box in c shows the location of the data shown in a and b. d, Partial P-wave stack. The boxes in d show the location of lower AMC amplitudes compared to that of the ~0–3 km stack (c) suggesting the presence of melt within the chamber. We note the strong AMC

image within the 2–3 km offset partial stack d, which indicates mush beneath the rest of the section. We interpret the melt zone by the lack of an AMC reflection in the partial P-wave stack, and presence of $P_{\text{melt}}S$ in the partial S-wave stack; and vice versa for mush zones a and b. e, Schematic diagram showing the locations of the melt and mush along the ridge axis. (Exact location of the pure melt sections are available; see Fig. A2 in Supplementary Information.)

reflection does not vary significantly along-axis⁹, so we use a waveform inversion method that assumes that the layers are horizontally stratified. The inversion scheme is implemented in intercept time-slowness domain²³. This transformation also allows a clear representation of the AMC and converted arrival ($P_{\text{melt}}S$) to be seen (Fig. 3) without interference of slow-phase-velocity events such as the sea-floor reflection. The waveform inversion method determines the P- and S-wave velocities of the crust by improving the fit between observed data and synthetically calculated data. It is an automated method, and therefore reduces human bias and also provides error estimates on the final solution^{6,23}. The large-scale initial P-wave velocity was taken from Tolstoy *et al.*⁹. The initial S-wave velocity was obtained from the P-wave velocity assuming a Poisson's ratio of 0.26 (ref. 24). We used a P-wave attenuation coefficient of 16 for the first 200 m, and 40–90 for the crust below²⁴. The S-wave attenuation coefficient was half of the P-wave attenuation coefficient. The model consists of a stack of 8-m-thick layers, for which the P-wave and S-wave velocities, density and attenuation are defined. This sampling interval was based on the minimum thickness that can be resolved from the observed data, which has a frequency bandwidth of 5–30 Hz. As data from all the slownesses were inverted simultaneously, the inverted results contain a model which is consistent with the data from all slownesses, and is therefore less likely to be influenced by incoherent noise due to two- and three-dimensional effects.

The turning rays above the sill and the P-wave reflection arrivals from the AMC constrain the P-wave and S-wave velocities above the AMC. Further constraint on S-wave velocity structure above the AMC comes from the arrival time of the $P_{\text{melt}}S$ phase. For this phase, the P-to-S conversion occurs at the solid–fluid interface of the AMC and at the sea bed such that the wave travels one leg between the sea bed and AMC as a P-wave, and the other as an S-wave. The waveforms of the AMC reflection and the $P_{\text{melt}}S$ phase constrain P-wave and S-wave velocities within the AMC and its neighbourhood. However, as with all seismic techniques, we have to guard against non-uniqueness of the final solution. To gain confidence in our results, we altered individual features within various candidate models while keeping the rest of the model fixed, and re-ran the inversion; the models that gave the best fit with the data are presented here.

Stacking. To enhance the partial-stack images, we firstly applied a frequency-wavenumber filter to the CMP gathers to remove sea-floor contamination of the AMC events, then performed a normal-moveout correction of 2.10 and 1.85 km s^{-1} for the P-wave and S-wave image, respectively. The partial stacks were obtained by stacking data between 2 and 3 km offsets, and all other processing parameters were as for the ~0–3 km offset stack. We note that the effect of two-dimensional scattering and multiple conversion points for $P_{\text{melt}}S$ will tend to provide a minimum estimate of the lateral extent of the pure melt zone. The absence of $P_{\text{melt}}S$ stack energy in the melt region could be due to the fact that the $P_{\text{melt}}S$ phase has two ray paths with slightly different points of conversion at the AMC that might affect the $P_{\text{melt}}S$ stack. Alternatively, the results could be confused by navigational error: that is, the ship track slipping off the axis in this region (Fig. 1).

Received 31 December 1997; accepted 28 May 1998.

- Detrick, R. S. *et al.* Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326**, 35–41 (1987).
- Detrick, R. S. *et al.* Seismic structure of the southern East Pacific Rise. *Science* **259**, 499–503 (1993).
- Kent, G. M., Harding, A. J. & Orcutt, J. A. Evidence for a smaller magma chamber beneath the East Pacific Rise at 9° 30' N. *Nature* **344**, 650–653 (1990).
- Mutter, J. C. *et al.* Magma distribution across ridge axis discontinuities on the East Pacific Rise from multi-channel seismic images. *Nature* **336**, 156–158 (1988).
- Hussenoeder, S. A. *et al.* Seismic analysis of the axial magma chamber reflector along the southern East Pacific Rise from conventional reflection profiling. *J. Geophys. Res.* **101**, 22087–22105 (1996).
- Collier, J. S. & Singh, S. C. Detailed structure of the top of the melt body beneath the East Pacific Rise at 9° 40' N from waveform inversion of seismic reflection data. *J. Geophys. Res.* **102**, 20287–20304 (1997).
- Kent, G. M., Harding, A. J. & Orcutt, J. A. Distribution of magma beneath the East Pacific Rise between the Clipperton transform and the 9° 17' Deval from forward modelling of common depth point data. *J. Geophys. Res.* **98**, 13945–13970 (1993).
- Kent, G. M., Harding, A. J. & Orcutt, J. A. Uniform accretion of oceanic crust south of the Garrett transform at 14° 15' S on the East Pacific Rise. *J. Geophys. Res.* **99**, 9097–9116 (1994).
- Tolstoy, M. *et al.* Deepening of the axial magma chamber on the southern East Pacific Rise towards the Garrett Fracture Zone. *J. Geophys. Res.* **102**, 3097–3108 (1997).
- Schmeling, H. Numerical models on the influence of partial melt on elastic, anelastic and electrical properties of rocks: I Elasticity and anelasticity. *Phys. Earth. Planet. Inter.* **41**, 34–57 (1985).
- Murase, T. & McBirney, A. R. Properties of some common igneous rocks and their melts at high temperature. *Geol. Soc. Am. Bull.* **84**, 3563–3592 (1973).
- Manghnani, M. H., Sato, H. & Rai, C. S. Ultrasonic velocity and attenuation measurements on basalt melts to 1500 °C: Role of composition and structure in the visco-elastic properties. *J. Geophys. Res.* **91**, 9333–9342 (1986).

- Mainprice, D. Modelling the anisotropic seismic properties of partially molten rocks found at mid-ocean ridges. *Tectonophysics* **279**, 161–179 (1997).
- Marsh, B. D. Magma chamber. *Annu. Rev. Earth Planet. Sci.* **17**, 439–474 (1989).
- Batiza, R. & Niu, Y. Petrology and magma chamber processes at the East Pacific Rise. *J. Geophys. Res.* **97**, 6779–6797 (1992).
- Hooff, E. E., Detrick, R. S. & Kent, G. M. Seismic structure and indicators of magma budget along the southern East Pacific Rise. *J. Geophys. Res.* **102**, 27319–27340 (1997).
- Collier, J. S. & Sinha, M. C. Seismic mapping of a magma chamber beneath the Valu Fa Ridge, Lau basin. *J. Geophys. Res.* **97**, 14031–14053 (1992).
- Macdonald, K. C. *et al.* A new view of the mid-ocean ridge from the behaviour of ridge-axis discontinuities. *Nature* **335**, 217–225 (1988).
- Sinton, J. M. *et al.* Magmatic process at superfast spreading mid-ocean ridges: Glass compositional variations along the East Pacific Rise 13°–23° S. *J. Geophys. Res.* **96**, 6133–6155 (1991).
- Langmuir, C. H., Bender, J. F. & Batiza, R. Petrological and tectonic segmentation of the East Pacific Rise. *Nature* **322**, 422–429 (1986).
- Worster, M. G., Huppert, H. E. & Sparks, R. S. Convection and crystallisation in magma cooled from above. *Earth Planet. Sci. Lett.* **101**, 79–89 (1990).
- Scott, D. R. & Stevenson, D. J. A self-consistent model of melting, magma migration and buoyancy-driven circulation beneath mid-ocean ridges. *J. Geophys. Res.* **94**, 2973–2988 (1989).
- Kormendi, F. & Dietrich, M. Non-linear waveform inversion of plane-wave seismograms in stratified elastic media. *Geophysics* **56**, 664–674 (1991).
- Christeson, G. L., Wilcock, W. S. D. & Purdy, G. M. The shallow attenuation structure of the fast-spreading East Pacific Rise near 9° 30' N. *Geophys. Res. Lett.* **21**, 321–324 (1994).
- Tarantola, A. *Inverse Problem Theory* (Elsevier, New York, 1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. This work started when one of us (S.C.S.) was a Cecil Green Scholar at Scripps Institution of Oceanography. J.S.C. is supported by a NERC (BRIDGE) fellowship. Critical reviews from R. Detrick, M. Sinha and J. Garmany significantly improved this Letter.

Correspondence and requests for materials should be addressed to S.C.S. (e-mail: singh@esc.cam.ac.uk).

The intensity of the Earth's magnetic field over the past 160 million years

M. T. Juárez*, L. Tauxe†, J. S. Gee† & T. Pick‡

* Fort Hoofddijk Paleomagnetic Laboratory, Budapestlaan 17, 3584 CD Utrecht, The Netherlands

† Scripps Institution of Oceanography, La Jolla, California 92093-0220, USA

‡ European Topic Center on Catalogue of Data Sources, Archivstrasse 2, D-30169 Hannover, Germany

In contrast to our detailed knowledge of the directional behaviour of the Earth's magnetic field during geological and historical times^{1,2}, data constraining the past intensity of the field remain relatively scarce. This is mainly due to the difficulty in obtaining reliable palaeointensity measurements, a problem that is intrinsic to the geological materials which record the Earth's magnetic field. Although the palaeointensity database has grown modestly over recent years^{3–5}, these data are restricted to a few geographical locations and more than one-third of the data record the field over only the past 5 Myr—the most recent database⁵ covering the time interval from 5 to 160 Myr contains only about 100 palaeointensity measurements. Here we present 21 new data points from the interval 5–160 Myr obtained from submarine basalt glasses collected from locations throughout the world's oceans. Whereas previous estimates for the average dipole moment were comparable to that of the Earth's present field⁶, the new data suggest an average dipole moment of $(4.2 \pm 2.3) \times 10^{22} \text{ A m}^2$, or approximately half the present magnetic-field intensity. This lower average value should provide an important constraint for future efforts to model the convective processes in the Earth's core which have been responsible for generating the magnetic field.

To augment the palaeointensity database, we focused on submarine basaltic glass (SBG) obtained from 20 sites sampled by the Deep Sea Drilling Project (DSDP). In addition to the DSDP glasses, we also include results from SBG obtained from pillow lavas in the Troodos ophiolite on Cyprus. SBG has been shown to contain predominantly single-domain magnetite as the carrier of the rema-

ment magnetization^{7,8}, based on a detailed rock-magnetic analysis. These studies also indicate that remanence-carrying grains alter little during heating in the laboratory. Analysis of SBG obtained from a recent lava flow recovered the geomagnetic field at the site⁹. These factors suggest that SBG is a nearly ideal material for palaeointensity studies.

The results presented here are from locations in both hemispheres, and are distributed throughout the world's oceans (Table 1). As the DSDP results were obtained from sites with identified marine magnetic anomalies, their ages are reasonably well constrained (Table 1). Moreover, tectonic-plate reconstructions allow adequate palaeolatitudes to be estimated and we therefore can calculate virtual axial dipole moments (VADM) for samples spanning the past 160 Myr.

Our experimental design is based on the modified stepwise double heating method described in refs 10 and 11. Briefly, after measurement of the natural remanent magnetization (NRM), specimens are heated to a given temperature. They are then either cooled in zero field (to assess the decay of the NRM) or in an applied field (to assess the acquisition of partial-thermal remanent magnetization, p-TRM). After treatment at a given temperature step, an in-field step at a lower temperature can be repeated (a p-TRM check) to determine if the capacity to acquire thermal remanent magnetism (TRM) has remained constant or has changed because of chemical alteration of the samples at elevated temperatures.

As illustrated in Fig. 1, the behaviour of this new SBG sample collection is quite similar to that of SBG results described earlier⁷⁻⁹, regardless of age or geographical location. Although DSDP glass samples are typically unorientated, we calculate the direction of the characteristic component and associated maximum angle of deviation¹² from the zero-field steps as a check on the validity of the palaeointensity results.

Maximum blocking temperatures of the NRM for most of the samples are generally between 475 and 550 °C, in agreement with previous results obtained for Cretaceous and Holocene SBG^{7,8}. Regardless of the origin of remanence (whether a TRM or an early

chemical remanence), the fact that SBG recovered from recent lava flows has similar rock magnetic characteristics and recovers the Earth's field accurately⁹ lends support to our contention that SBG is a reliable recording medium of the Earth's palaeofield.

The wealth of data provided by the modified Thellier–Thellier technique as practised here allows ample information for objective data selection. Our criteria are as follows. (1) The component of magnetization used to determine palaeointensity must trend to the origin (within the maximum angle of deviation and the maximum angle of deviation must itself be <15°). (2) A best-fit line through the NRM/p-TRM data (see, for example, Fig. 1c and d) must have low scatter based on standard error of the slope (see, for example, ref. 11). (3) The p-TRM checks (triangles in Fig. 1c and d) must agree with the original p-TRM within 5%. (4) Data from multiple specimens of the same glassy margin must agree within 10%. These were then averaged to give a single palaeointensity estimate.

Data from specimens meeting the above criteria were deemed acceptable for palaeointensity calculations. Complete information concerning specimens used here is available: see Supplementary Information. Averages by hole are listed in Table 1.

The palaeointensity values from a single location are relatively uniform with the exception of those from Hole 238 (see Table 1) where two populations of intensity data can be statistically distinguished; the upper cores show higher values and are more scattered than the lower cores (Table 1). We have therefore split these data into two groups (upper and lower), and treat them as independent estimates of the geomagnetic field.

The absolute value of the slope of the best-fit line through the NRM/p-TRM data (see, for example, Fig. 1c and d) multiplied by the value of the magnetic field in the laboratory gives the absolute palaeofield (*B*) of the specimen in microteslas. These can be converted to VADM, given the palaeolatitudes which were estimated using appropriate palaeomagnetic data¹³⁻¹⁵ or plate reconstructions¹⁶⁻¹⁸.

In Fig. 2 we plot VADM values for all samples of SBG available, including those previously published^{7,9,19}. The VADM values for the

Table 1 Locations, ages, remanence and palaeointensity data for samples plotted in Fig. 2.

Hole*	Plate†	MMA‡	Age (Myr)§	λ/ϕ	λ' ¶	<i>B</i> (s/n)#	VADM°	<i>M_r/M_{eq}</i> **
504B	NZ	3A	6.4 ± 0.2	1/-84	1	30.3 ± 5.1 (3/3)	7.8 ± 1.3	6.6/6.6
395A	AF	4n	8 ± 1	23/-46	23	12.7 ± 8.2 (17/11)	2.6 ± 1.8	3.4/2.8
410/410A	NA	5n	10.3 ± 0.7	46/-30	45	10.8 ± 0.3 (4/3)	1.8 ± 0.1	3.2/2.0
396B	AF	5n	11.5 ± 0.5	23/-44	23	9.2 ± 2.2 (7/5)	2.0 ± 0.5	1.8/1.5
335	NA	5A-5B	13.8 ± 1.3	37/-35	37	16.2 ± 6.4 (10/9)	2.9 ± 1.1	3.7/2.6
520	AF	5Br	14.9 ± 0.2	-26/-11	-26	29.0 (2/1)	5.9	7.8/6.2
470A	NA	5B	15.7 ± 0.7	29/-118	24	31.7 ± 7.6 (3/3)	6.6 ± 1.6	5.3/4.3
562	NA	5Dr	17.9 ± 0.4	33/-42	32	13.6 ± 3.4 (9/9)	2.6 ± 0.6	3.3/2.4
495	CO	6	22 ± 2	12.5/-91	2	17.3 ± 2.1 (6/4)	4.4 ± 0.5	0.8/0.8
556	NA	12	32 ± 1.1	39/-35	36	15.8 ± 4.9 (5/4)	2.9 ± 0.9	3.3/2.3
559	NA	12r	32 ± 1.1	35/-41	32	9.5 ± 0.7 (3/2)	1.8 ± 0.1	0.5/0.4
563/564	NA	13	33.3 ± 0.3	34/-44	31	16.3 ± 3.2 (13/9)	3.1 ± 0.6	3.2/2.4
238 (upper)	IN	11 - 13	34 ± 4	-11/71	-24	33.0 ± 9.6 (5/3)	6.9 ± 2.0	1.7/1.4
238 (lower)	IN	11 - 13	34 ± 4	-11/71	-24	9.4 ± 0.5 (7/5)	2.0 ± 0.1	
407	NA	15r	35.2 ± 0.3	64/-31	60	27.9 ± 3.6 (5/4)	4.0 ± 0.5	5.0/2.8
522B	AF	16	35.7 ± 1.3	-27/-5	-30	19.6 ± 3.2 (6/4)	3.8 ± 0.6	5.0/3.8
447A	PH	18 - 21	43.5 ± 5.5	18/133	10	33.5 ± 0.7 (2/2)	8.3 ± 0.2	3.6/3.4
525A	AF	32n	72 ± 1	-29/3	-38	51.3 ± 3.9 (10/5)	9.0 ± 0.7	14.0/9.6
543A††	NA	34(y)	84	16/-60	-21	17.9 ± 2.2 (12/11)	2.0 ± 0.6	7.4/7.1
TR	AF	CNS	92	35/33	20	36.0 ± 6.9 (6/5)	7.9 ± 1.5	7.9/6.8
462A	NU	CNS	111 ± 1	7/165	-7	14.8 ± 0.4 (3/2)	3.7 ± 0.1	9.2/9.0
807C††	PA	CNS	122.3 ± 1	4/157	-34	19.4 ± 7.4 (30/18)	3.6 ± 1.4	5.9/4.2
417D††	NA	M0	125 ± 1	25/-68	17	12.0 ± 3 (12/11)	2.8 ± 0.7	9.2/8.2
418A††	NA	M0	125 ± 1	25/-68	17	15.5 ± 9 (12/11)	3.5 ± 1.9	13.6/12.1
534A	NA	M28	156.5 ± 0.1	28/-75	9	12.0 (1/1)	3.0	3.6/3.5

* DSDP/ODP hole, TR is Troodos ophiolite.

† Tectonic plate: AF, African; CO, Cocos; IN, Indian; NA, North American; NZ, Nazca; PH, Philippine; NU, Nauru.

‡ Marine magnetic anomaly (MMA) measured above site, CNS is the Cretaceous normal superchron.

§ Refs 33-36.

|| Latitude (λ), and longitude (ϕ) of site in degrees.

¶ Estimated palaeolatitude in degrees (λ' , see text).

Field intensity in μ T (number of samples measured/number of samples used).

° Virtual axial dipole moment (10^{22} A m²).

** NRM intensity and equatorial equivalent, in A m⁻¹. For sources, see Supplementary Information.

†† Data published earlier⁹ and included for completeness.

pre-5-Myr data (Table 1) range from a minimum of $1.8 \times 10^{22} \text{ A m}^2$ for Sites 410 and 559 (10 and 32 Myr, respectively) to a maximum of $9.0 \times 10^{22} \text{ A m}^2$ for Site 525A (72 Myr). The average dipole moment calculated from these new data is $(4.2 \pm 2) \times 10^{22} \text{ A m}^2$, approximately half that of the present-day field. Furthermore, the highest values come from Late Cretaceous times (Hole 525A, and the Troodos ophiolite), a period with an anomalously low rate of reversals.

For comparison, we plot the palaeointensity database⁵ as open circles in Fig. 2. Our only selection criteria for these data was that they be treated by a "Thellier" type of experiment, that they be "non-transitional" and that they have more than one specimen per site. There are 164 data points meeting these minimal criteria with ages between 0 and 160 Myr. In general, there is a fair degree of correspondence between the database and our new data set, although we find no high values in the time period around 15 Myr. Most of these high values come from the Steens Mountain

study²⁰, a study with good temporal resolution. A single DSDP/ODP hole cannot hope to achieve the temporal resolution of such a stratigraphic profile, sampling instead at discrete intervals. It is therefore likely that the high values of the field evident in the Steens Mountain collection²⁰ were missed in our sample collection.

The remanent magnetization of the oceanic crust appears to undergo systematic variations over time²¹. We have compiled remanence values, M_{eq} (adjusted for the effects of palaeolatitude to equivalent equatorial values), for the dredge collections used in ref. 7, the Troodos pillow lavas, and DSDP basalts (Fig. 2). Data from DSDP/ODP holes were dated and adjusted for palaeolatitude in the same way as the palaeointensity data discussed above. The data that come from the same holes as our palaeointensity data are listed in Table 1. The complete data set is available: see Supplementary Information.

As illustrated in Fig. 2, ridge-crest basalts range in M_{eq} values

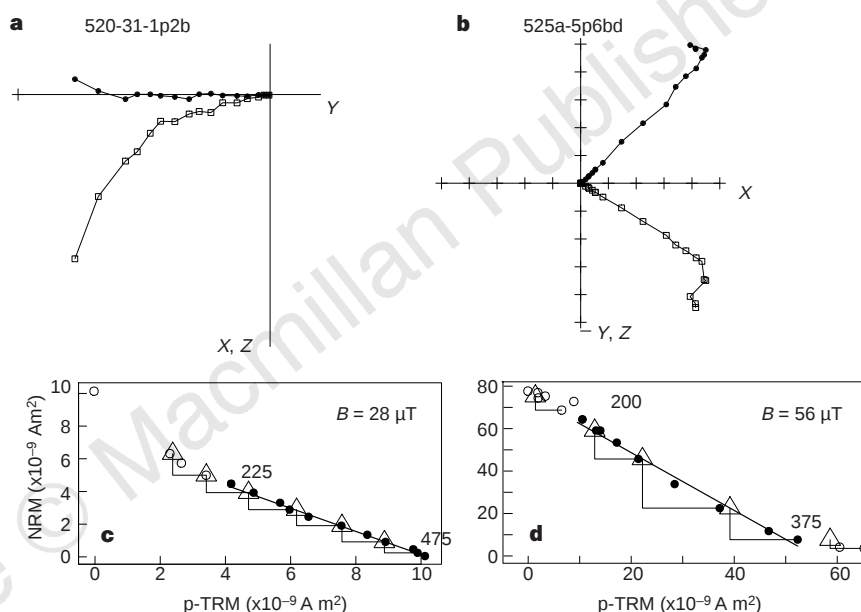


Figure 1 Behaviour of magnetic remanence during the Thellier-Thellier experiments. **a** and **b**, Vector end-point diagrams for the zero-field steps from 15-Myr and 72-Myr samples. Filled symbols are horizontal projections and open symbols are vertical projections. (Samples are unorientated). Treatment steps were: NRM, 100 °C, 150 °C then in 25 °C intervals to maximum. **c** and **d**, Plot of NRM

remaining (zero-field steps) versus p-TRM gained (in field steps) from same samples shown in **a** and **b**. Filled symbols are used in the palaeofield calculation. Open triangles are the repeated p-TRM steps after heating to a higher temperature.

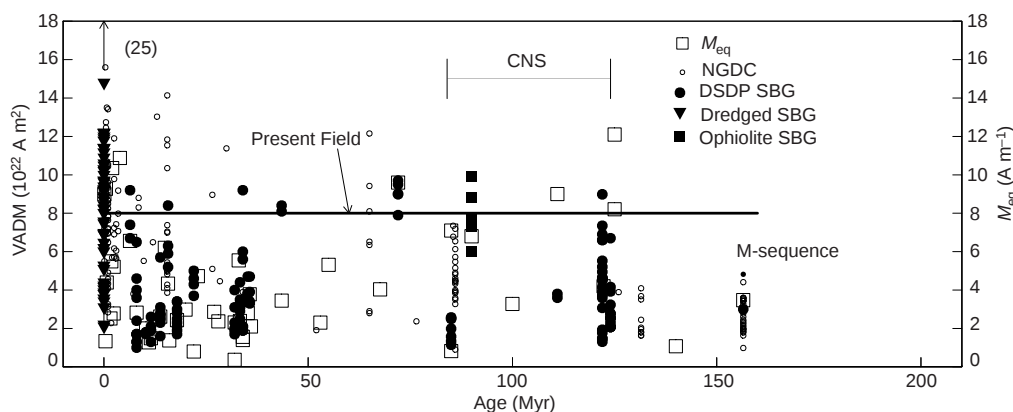


Figure 2 Summary of palaeointensity data for the past 160 Myr and NRM data from the DSDP/ODP basalts. Filled squares, estimated VADMs for palaeointensity from submarine basaltic glass (SBG); small circles, VADMs from the Thellier-

Thellier database at the National Geophysical Data Center (NGDC); open squares, remanent magnetization (reduced to the equator, M_{eq}) of DSDP/ODP basalts versus age. Values of zero-age magnetization from ref. 32.

from less than 1 A m^{-1} to over 25 A m^{-1} at the ridge crest. With age, these values decrease rapidly to an average of several A m^{-1} . Values around 3 A m^{-1} are typical of DSDP basalts throughout the Neogene period. In the Late Cretaceous, high values of $\sim 10 \text{ A m}^{-1}$ are encountered, corresponding to the Cretaceous normal superchron (CNS) and a few million years after the end of the CNS. The period of time corresponding to the M-sequence anomalies (124–160 Myr) is characterized by low values of M_{eq} (and also by the “Mesozoic dipole low”²²).

The general shape of the M_{eq} curve was first described by Bleil and Petersen²¹. The natural remanence of oceanic crustal rocks is controlled by a number of factors. Thermal remanence is controlled to first order by the intensity of the magnetic field and the amount and rock-magnetic characteristics of the magnetizable material in the rock. Even the youngest sea-floor basalts have apparently experienced some degree of alteration^{23,24}, and the exact behaviour of remanence during alteration is a complicated matter²⁵. Thus, the long-term behaviour of the M_{eq} curve shown in Fig. 2 could result from (1) changes in palaeofield intensity, (2) changes in bulk chemistry (for example, Fe content), and/or (3) alteration history. Bleil and Petersen²¹ argued that the behaviour of remanent intensities was primarily the result of gradual alteration of magnetic minerals of the oceanic crust by a mechanism of cation diffusion. Raymond and LaBrecque²⁶ proposed a quantitative alteration model in which a chemical remanent magnetization (CRM) is acquired as the original TRM is lost through low-temperature oxidation. According to their model, CRM acquired over the past 20 Myr could account for 80% of the variation in M_{eq} . Johnson and Pariso²⁷ argued that the cause of the observed variation of NRM in oceanic basalts was controlled primarily by differences in the original content of FeTi oxides in these older rocks. Thus, it would appear that the control of the long-term variation of the NRM of the ocean crust is still a matter of debate.

Although more difficult to demonstrate, some studies have suggested that a record of past geomagnetic intensity variations might be preserved in the magnetization of the ocean crust^{28–30}. Several earlier investigations discounted the hypothesis that the intensity of the Earth’s magnetic field played a significant role in the variations of the NRM on the basis of the lack of evidence for similar variations in the palaeointensity databases of the time. However, the new data set shown in Fig. 2 provides support for the view that the long-term variations in NRM of sea-floor basalts may, in large part, be a function of palaeofield intensity.

A last point that needs to be discussed concerns the effect that alteration of the magnetic minerals could have on the palaeointensity determinations. It is not yet completely clear whether our SBGs have undergone alteration that may have influenced their ability to record past geomagnetic intensity fluctuations. We deliberately select samples that have a ‘fresh’ appearance, excluding devitrified or evidently altered material. Although differentiating between TRM and CRM is still extremely difficult, the similarity in behaviour throughout our data set argues against gradual alteration. Because the youngest samples of SBG were collected within one year of eruption⁹, any alteration that took place occurred within the first year, and produced a remanence that was indistinguishable from the original TRM.

Three conclusions can be drawn from our study. (1) Our data suggest that the average palaeofield intensity for the period 0–160 Myr is $4 \pm 2 (10^{22} \text{ A m}^2)$. These values are similar to the “Mesozoic dipole low” of Prévot *et al.*²², suggesting that the Mesozoic field was not “low” but of average intensity. As has been previously suggested on the basis of analysis of marine magnetic anomalies²⁹, it appears that the present field may be anomalously high with respect to the long-term average dipole intensity. (2) In the collection of palaeointensity data from SBG, high palaeofield values are obtained for the most recent field⁷ (the past 3 kyr) and for

the Late Cretaceous (Table 1). The Late Cretaceous is characterized by a much lower rate of reversals than during the M-sequence or the Cenozoic era. The correspondence of higher field values with lower reversal rate is consistent with the data set of Tauxe and Hartl³¹, who presented a correlation of relative palaeointensity data with polarity interval length. (3) To first order, the long-term trend in palaeointensity data from submarine basaltic glass parallels the trends in remanence behaviour from DSDP basalts. This suggests that the remanence behaviour is controlled, at least in part, by geomagnetic field intensity variations. □

Received 2 March; accepted 3 June 1998.

- Opdyke, N. D. & Channell, J. E. T. *Magnetic Stratigraphy* (Academic, San Diego, 1996).
- Merrill, R. T., McElhinny, M. W. & McFadden, P. L. *Paleomagnetism, the Core, and the Deep Mantle* (Academic, San Diego, 1996).
- McFadden, P. L. & McElhinny, M. W. Variations in the geomagnetic dipole 2: statistical analysis of VDM’s for the past 5 m.y. *J. Geomagn. Geoelectr.* **34**, 163–189 (1982).
- Tanaka, H., Kono, M. & Uchimura, H. Some global features of paleointensity in geological time. *Geophys. J. Int.* **120**, 97–102 (1995).
- Perrin, M. & Shcherbakov, V. Paleointensity of the Earth’s magnetic field for the past 400 Ma: evidence for a dipole structure during the Mesozoic low. *J. Geomagn. Geoelectr.* **49**, 601–614 (1997); data available by anonymous ftp at (<ftp://ftp.ngdc.noaa.gov>) (cited 17 July 1998).
- Merrill, R. T. & McElhinny, M. W. *The Earth’s Magnetic Field* (Academic, San Diego, 1983).
- Pick, T. & Tauxe, L. Holocene paleointensities: Thellier experiments on submarine basaltic glass from the East Pacific Rise. *J. Geophys. Res.* **98**, 17949–17964 (1993).
- Pick, T. & Tauxe, L. Characteristics of magnetite in submarine basaltic glass. *Geophys. J. Int.* **119**, 116–128 (1994).
- Pick, T. & Tauxe, L. Geomagnetic paleointensities during the Cretaceous normal superchron measured using submarine basaltic glass. *Nature* **366**, 129–139 (1993).
- Thellier, E. & Thellier, O. Sur l’intensité du champ magnétique terrestre dans le passé historique et géologique. *Ann. Geophys.* **15**, 285–378 (1959).
- Coe, R. S., Grommé, S. & Mankinen, E. A. Geomagnetic paleointensities from radiocarbon-dated lava flows on Hawaii and the question of the Pacific nondipole low. *J. Geophys. Res.* **83**, 1740–1756 (1978).
- Kirschvink, J. L. The least-squares line and plane and the analysis of paleomagnetic data. *Geophys. J. R. Astron. Soc.* **62**, 699–718 (1980).
- Van der Voo, R. *Paleomagnetism of the Atlantic, Tethys and Iapetus Oceans* (Cambridge Univ. Press, 1993).
- Steiner, M. B. Paleomagnetism of the Cretaceous section, Site 462. *Init. Rep. DSDP* **61**, 711–716 (1981).
- Clube, T. M. M., Creer, K. M. & Robertson, A. H. F. Paleorotation of the Troodos microplate, Cyprus. *Nature* **317**, 522–525 (1985).
- Mueller, R. D., Royer, J.-Y. & Lawver, L. A. Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* **21**, 275–278 (1993).
- Lee, T.-Y. & Lawver, L. A. Cenozoic plate reconstruction of Southeast Asia. *Tectonophysics* **251**, 85–138 (1995).
- Gripp, A. & Gordon, R. Plate velocities relative to hot spots. *Geophys. Res. Lett.* **17**, 1109–1112 (1990).
- Mejia, V., Opdyke, N. D. & Perfit, M. R. Paleomagnetic field intensity recorded in submarine basaltic glass from the East Pacific Rise, the last 69 ka. *Geophys. Res. Lett.* **23**, 475–478 (1996).
- Prévot, M., Mankinen, E. A., Coe, R. S. & Grommé, C. S. The Steens Mountain (Oregon) geomagnetic polarity transition 2. Field intensity variations and discussion of reversal models. *J. Geophys. Res.* **90**, 10417–10448 (1985).
- Bleil, U. & Petersen, N. Variations in magnetization intensity and low-temperature titanomagnetite oxidation of ocean floor basalts. *Nature* **301**, 384–388 (1983).
- Prévot, M., Derder, M. E. M., McWilliams, M. & Thompson, J. Intensity of the Earth’s magnetic field: evidence for a Mesozoic dipole low. *Earth Planet. Sci. Lett.* **97**, 129–139 (1990).
- Johnson, H. P. & Trivey, M. A. Magnetic properties of zero-age oceanic crust; a new submarine lava flow on the Juan de Fuca ridge. *Geophys. Res. Lett.* **22**, 175–178 (1995).
- Kent, D. V. & Gee, J. Magnetic alteration of zero-age oceanic basalt. *Geology* **24**, 703–706 (1996).
- Dunlop, D. J. & Özdemir, Ö. *Rock Magnetism: Fundamentals and Frontiers* (Cambridge Univ. Press, 1997).
- Raymond, C. & LaBrecque, J. L. Magnetization of the oceanic crust: thermoremanent magnetization or chemical remanent magnetization? *J. Geophys. Res.* **92**, 8077–8088 (1987).
- Johnson, H. P. & Pariso, J. E. Variations in oceanic crustal magnetization: Systematic changes in the last 160 million years. *J. Geophys. Res.* **98**, 435–445 (1994).
- Cande, S. C. & Kent, D. V. Ultrahigh resolution marine magnetic anomaly profiles: a record of continuous paleointensity variations? *J. Geophys. Res.* **97**, 15075–15083 (1992).
- Gee, J., Schneider, D. A. & Kent, D. V. Marine magnetic anomalies as recorders of geomagnetic intensity variations. *Earth Planet. Sci. Lett.* **144**, 327–335 (1996).
- Johnson, H. P., Vanpatten, D. & Sager, W. W. Age-dependent variation in the magnetization of seamounts. *Geophys. Res. Lett.* **22**, 231–234 (1995).
- Tauxe, L. & Hartl, P. 11 million years of Oligocene geomagnetic field behaviour. *Geophys. J. Int.* **128**, 217–229 (1996).
- Gee, J. & Kent, D. V. Magnetization of axial lavas from the southern East Pacific Rise (14 degrees–23 degrees S): Geochemical controls on magnetic properties. *J. Geophys. Res.* **10**, 24873–24886 (1997).
- Castillo, P. R., Pringle, M. S. & Carlson, R. W. East Mariana Basin tholeiites: Cretaceous intraplate basalts or rift basalts related to the Ontong Java plume? *Earth Planet. Sci. Lett.* **123**, 139–154 (1994).
- Mukasa & Ludden *Geology* **15**, 825 (1987).
- Harland, W. B. *et al.* *A Geologic Time Scale* (Cambridge Univ. Press, 1989).
- Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity timescale for the Late Cretaceous and Cenozoic. *J. Geophys. Res.* **100**, 6093–6095 (1995).

Supplementary information is available on Nature’s World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. This work was supported in part by the NSF and GOA.

Correspondence and requests for materials should be addressed to L.T. (e-mail: ltaxe@ucsd.edu).

The evolution of warning signals

Shigeo Yachi* & Masahiko Higashi†

* Laboratoire d'Ecologie, UMR 7625, Ecole Normale Supérieure, 46 rue d'Ulm, F-75230, Paris Cedex 05, France

† Center for Ecological Research, Kyoto University, Kyoto 606-01, Japan

Warning coloration signals are a familiar and conspicuous phenomenon in nature. However, the fundamental question of how warning signals initially evolved remains unanswered. For an unpalatable prey to evolve a signal to indicate its unprofitability, a rare and conspicuous mutant in a population of unpalatable cryptic prey must overcome a double disadvantage: a greater risk of being detected (as a result of being more conspicuous) and of being attacked (because its rarity results in a decreased association with aversion) by a predator^{1,2}. Although the prior evolution of prey gregariousness may help warning signals to evolve^{3–8}, such an evolutionary order may not always be the case^{4,9–11}. Here we present a theoretical model that describes a mechanism for the evolution of warning signals without having to invoke gregariousness. Specifically, a predator's generalization of stimulus in associative learning, with a 'peak shift' towards greater conspicuousness^{5,12–15}, allows a warning signal to evolve when the prey population density exceeds a threshold. Once a warning signal starts to evolve, it continues to grow; the resulting, evolutionarily stable¹⁶ conspicuousness of prey is discontinuously greater than that of the original cryptic prey, drawing an unambiguous distinction in their appearance.

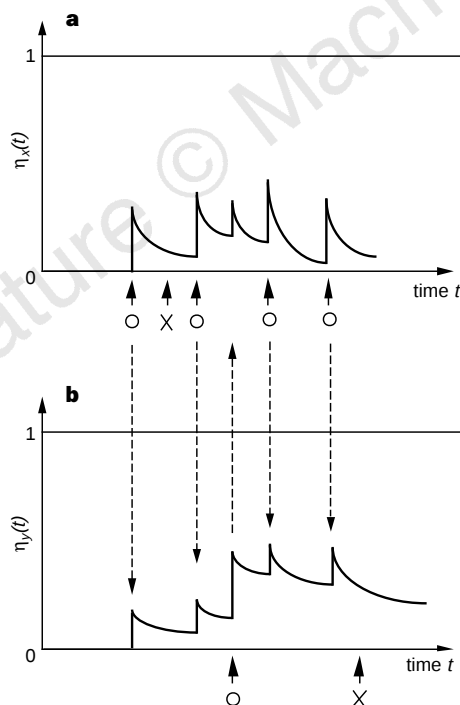


Figure 1 The stochastic change in the level of associative memories of a predator who randomly searches a prey population consisting of two prey types. For prey types x and y , the predator's associative memory is given by **a**, $\eta_x(t)$ and **b**, $\eta_y(t)$, respectively. When the predator detects a prey individual (indicated by the arrows), it avoids (crosses) or attacks (circles) it. When the predator attacks and samples one type of prey, its associative memory for that type increases, whereas that for the other prey type also increases (the latter 'indirect learning' is indicated by dotted lines with an arrow), and thereafter both decay exponentially. Assume that $\eta_x(0) = \eta_y(0) = 0$.

Consider a prey population, randomly distributed in a habitat and consisting of two types (type x and type y) of individuals with the same degree of unpalatability, u (≥ 0), and different levels x and y ($0 \leq x < y$) of conspicuousness, and a predator that randomly forages on the prey. Let N_x and N_y , respectively, denote the population densities of type x and type y .

Let $\eta_x(t)$ denote the probability of recall of an unpleasant experience by the predator when it detects a prey individual of type x at time t , causing avoidance of the prey individual. Thus, $\eta_x(t)$ represents the predator's level of associative memory for type x prey at time t . The predator therefore attacks a detected prey individual of type x with probability $1 - \eta_x(t)$. When the predator samples (that is, attacks) a prey individual of type x , its associative memory $\eta_x(t)$ increases by $a(x, x, u)$, where $0 \leq a(x, x, u) \leq 1$ and $a(x, x, u)$ is a non-decreasing function of x and u , as a result of the unpleasant sampling^{17,18}. The associative memory $\eta_x(t)$ for type x prey also increases when the predator samples a prey individual of type y , but by a different amount $a(x, y, u)$, as long as y is sufficiently close to x (that is, indirect associative learning through generalization^{5,12–14}). Between such increment events, the associative memory $\eta_x(t)$ decays exponentially¹⁹ with a constant rate c and we will assume $c = 1$ for simplicity. The predator's associative memory for type y prey, $\eta_y(t)$, changes stochastically in a similar way (Fig. 1).

The fitness component of type x prey relevant to the evolution of aposematism (warning signals) is the prey's lifetime survival probability under predation $S_x = \exp[-\int_0^T p(x)(1 - \eta_x(t))\varphi(u)dt]$, where T , $p(x)$ and $\varphi(u)$ denote, respectively, the prey's lifespan, the predator's efficiency in detecting type x prey (assumed to be an

Box 1 Derivation of condition (1)

The fitness component of type x prey that is relevant to the evolution of aposematism can be expressed as

$$S_x = \exp\left[-\int_0^T p(x)(1 - \eta_x(t))\varphi(u)dt\right] = \exp[-p(x)(1 - \eta_x)\varphi(u)T] \quad (4)$$

where $\eta_x = \frac{1}{T} \int_0^T \eta_x(t)dt$. Likewise, the fitness component of type y prey is given as

$$S_y = \exp[-p(y)(1 - \eta_y)\varphi(u)T] \quad (5)$$

where $\eta_y = \frac{1}{T} \int_0^T \eta_y(t)dt$. Using shot-noise process²² approximation, the predator's average levels of associative memory η_x and η_y can be approximated as follows (S.Y. and M.H., in preparation):

$$\eta_x = \frac{(AD - BC) + (A + B)}{(AD - BC) + A + D + 1} \quad (6)$$

$$\eta_y = \frac{(AD - BC) + (C + D)}{(AD - BC) + A + D + 1} \quad (7)$$

where $A = p(x)N_x a(x, x, u)/c(x, u)$, $B = p(y)N_y a(x, y, u)/c(x, u)$, $C = p(x)N_x a(y, x, u)/c(y, u)$, $D = p(y)N_y a(y, y, u)/c(y, u)$ and we assume that the memory-decay constant c is in general a function of x and u , that is, $c(x, u)$. The criterion for a more conspicuous mutant y to dominate wild-type x , $S_x < S_y$, can, from equations (4) and (5), be written as: $p(x)(1 - \eta_x) > p(y)(1 - \eta_y)$, where $p(x) < p(y)$ because $x < y$. This is, from equations (6) and (7), equivalent to

$$\frac{1 + A - C}{1 - B + D} \frac{p(y)}{p(x)} < 1$$

or

$$\frac{[1/p(x)] + N_x[a(x, x, u)/c(x, u) - a(y, x, u)/c(y, u)]}{[1/p(y)] + N_y[a(y, y, u)/c(y, u) - a(x, y, u)/c(x, u)]} < 1 \quad (8)$$

When mutant y occurs by mutation in a major gene, such that generalization does not occur between x and y (that is, $a(x, y, u) = a(y, x, u) = 0$), this criterion (equation (8)) is reduced to

$$\frac{[1/p(x)] + N_x a(x, x, u)/c(x, u)}{[1/p(y)] + N_y a(y, y, u)/c(y, u)} < 1$$

which is further reduced to condition (1) when $c(x, u) = c(y, u) = 1$.

increasing function of x), and the probability with which a prey individual when attacked will be harmed such that it dies. The fitness S_y of type y prey is given similarly.

Suppose mutant y occurs as a result of a mutation in a major gene, causing the mutant patterns to deviate so greatly from wild-type x that generalization does not occur between them (that is, $a(x, y, u) = a(y, x, u) = 0$). The criterion for a more conspicuous mutant, y , to dominate the wild-type x (and therefore increase in frequency), $S_x < S_y$, would then become (Box 1):

$$N_y > \frac{1}{a(y, y, u)} \left[\left(\frac{1}{p(x)} - \frac{1}{p(y)} \right) + a(x, x, u)N_x \right] \quad (1)$$

This implies that, for a more conspicuous mutant to dominate, there exists a threshold to be exceeded by N_y , the mutant density^{1,2,6,20,21}, which increases with an increase in the wild-type density, N_x . Thus, it indicates that the rarity of a more conspicuous mutant ($N_y \ll N_x$) hinders the mutant from invading, which confirms the second barrier to the initial evolution of warning signals mentioned earlier^{1,2}.

However, when mutation occurs by minor genes, generalization is expected to occur, and the dominance criterion of the mutant becomes (Box 2):

$$Np(x)[\partial a(y, x, u)/\partial y]_{y=x} > \frac{p'(x)}{p(x)} \quad (2)$$

which represents the condition for a slightly more conspicuous

Box 2 Derivation of condition (2)

The criterion shown in equation (8) can be rewritten as:

$$N_x \left(\frac{a(y, x, u)}{c(y, u)} - \frac{a(x, x, u)}{c(x, u)} \right) + N_y \left(\frac{a(y, y, u)}{c(y, u)} - \frac{a(x, y, u)}{c(x, u)} \right) > - \left(\frac{1}{p(y)} - \frac{1}{p(x)} \right) \quad (9)$$

Because we assume that mutants occur as a result of mutations in minor genes, that is, $y = x + dx$ ($0 < dx \ll 1$), using the Taylor expansion and neglecting the terms higher than the first order, condition (9) can be simplified as follows. The right-hand side of equation (9) can be written as

$$- \left(\frac{1}{p(y)} - \frac{1}{p(x)} \right) = \frac{p'(x)}{p(x)^2} dx$$

The first term of the left-hand side of equation (9) can (by the definition of partial derivative) be written as

$$N_x \left(\frac{a(y, x, u)}{c(y, u)} - \frac{a(x, x, u)}{c(x, u)} \right) = N_x \left[\frac{\partial(a(y, x, u)/c(y, u))}{\partial y} \right]_{y=x} dx$$

whereas the second term can be rewritten as

$$\begin{aligned} N_y \left(\frac{a(y, y, u)}{c(y, u)} - \frac{a(x, y, u)}{c(x, u)} \right) &= N_y \left[\frac{a(x, x, u)}{c(x, u)} + \frac{(a_1(x, x, u) + a_2(x, x, u))c(x, u) - a(x, x, u)c'(x, u)}{c(x, u)^2} dx \right. \\ &\quad \left. - \frac{a(x, x, u) + a_2(x, x, u)dx}{c(x, u)} \right] \\ &= N_y \frac{a_1(x, x, u)c(x, u) - a(x, x, u)c'(x, u)}{c(x, u)^2} dx \\ &= N_y \left[\frac{\partial(a(y, x, u)/c(y, u))}{\partial y} \right]_{y=x} dx \end{aligned}$$

where $a_1(x, x, u) = [\partial a(y, x, u)/\partial y]_{y=x}$ and $a_2(x, x, u) = [\partial^2 a(y, x, u)/\partial y^2]_{y=x}$. Thus, equation (9) can in approximation be written as

$$N_x p(x) \left[\frac{\partial(a(y, x, u)/c(y, u))}{\partial y} \right]_{y=x} + N_y p(x) \left[\frac{\partial(a(y, x, u)/c(y, u))}{\partial y} \right]_{y=x} > \frac{p'(x)}{p(x)}$$

which, because $N_x + N_y = N$, implies the following:

$$Np(x)[\partial a(y, x, u)/\partial y]_{y=x} > p'(x)/p(x) \quad (10)$$

This condition is reduced to condition (2) when $c(y, u) = 1$.

mutant to invade and fix, noting that the mutant should fix if it can invade because condition (2) does not depend on its frequency.

For condition (2) to hold, it is necessary that $[\partial a(y, x, u)/\partial y]_{y=x} > 0$. This inequality implies that the associative learning (the increment $a(y, x, u)$ in associative memory) due to sampling a prey individual with some conspicuousness x , attains its maximum at a greater conspicuousness y (rather than at the sampled conspicuousness x itself). One well known mechanism to generate this shift of the maximum response in a predator's generalization is by a 'peak shift' through the predator's simultaneous discrimination learning with other palatable and cryptic species^{5,12-15}. Here we use the term 'peak shift' in a broader sense to represent the property of a predator's generalization shifting the maximum response towards greater conspicuousness, regardless of the mechanisms used to attain it.

Condition (2) shows that a larger population density N , which, at the time of the mutant's invasion, mostly consists of the wild type, facilitates the initial evolution of aposematism. This is in contrast to the case of a mutant by a major gene, in which a larger population density of the wild type puts a higher barrier for the initial evolution of aposematism. This contrast shows that generalization with a peak-shift property, which is effective only in the case of small mutations, is the very mechanism that resolves the second barrier to the initial evolution of aposematism.

Condition (2) also indicates that the tendency for a more conspicuous mutant to be more easily detected by the predator ($p'(x) > 0$) hinders a more conspicuous mutant from invading, which confirms the first barrier to the initial evolution of aposematism mentioned earlier^{1,2}, and that the peak shift ($[\partial a(y, x, u)/\partial y]_{y=x} > 0$) may, combined with a large population density N , overcome this barrier ($p'(x) > 0$) to bring about the evolution of aposematism.

Given a peak shift, that is $[\partial a(y, x, u)/\partial y]_{y=x} > 0$, condition (2) can

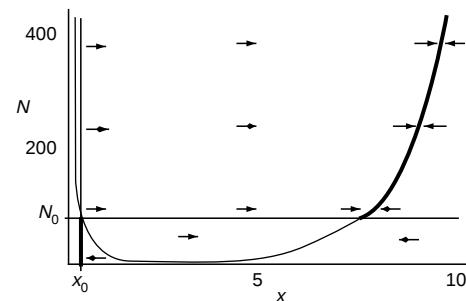


Figure 2 The evolutionary dynamics of the degree x of conspicuousness of the prey population for the case in which $a(y, x, u) = u(0.1 + x)/(1 + x) \times \exp(-x - y)^2(1 - 0.5 \exp(-y))$, $u = 1$ and $p(x) = x/(1 + x)$. We assume $p(0) = 0$, which means that $x = 0$ defines the state that achieves the complete escape from the predator's detection. Therefore, the original conspicuousness x of the cryptic wild type is expected to take some positive value $x_0 > 0$.

Box 3 Technical notes on condition (3)

The assumption that $p'(x) > 0$ and $0 \leq p(x) \leq 1$ implies that $p(x)$ converges to a finite value, and thus $p'(x)$ converges to 0, as x goes to ∞ . Assume that the degree of peak shift, represented by $a_1(x, x, u) = [\partial a(y, x, u)/\partial y]_{y=x}$, converges to 0 more quickly than $p'(x)$ does. Then, the right-hand side of condition (3) increases to infinity (because $p'(x)/a_1(x, x, u)$ does) as x goes to ∞ and (because $1/p(x)$ does) as x goes to 0. Thus, the curve line corresponding to inequality (3) in general takes a form such as the one in Fig. 2.

If the effect of u is separable in $a(y, x, u)$, that is, $a(y, x, u) = \alpha(y, x)h(u)$, condition (3) becomes $Nh(u) > p'(x)/p(x)^2 [\partial \alpha(y, x)/\partial y]_{y=x}$, which indicates that u has an effect similar to that of N on the evolution of aposematism.

Box 4 The effects of the conspicuousness dependence in memory retention

For condition (10) of aposematic evolution to hold, the derivative on its left-hand side must be positive, that is, the following inequality must hold:

$$\frac{1}{c(x,u)} \left[\frac{\partial a(y,x,u)}{\partial y} \right]_{y=x} + \frac{1}{c(x,u)^2} a(x,x,u) (-c_x(x,u)) > 0 \quad (11)$$

Thus, either $[\partial a(y,x,u)/\partial y]_{y=x} > 0$ or $c_x(x,u) < 0$ must hold. Whereas the first inequality implies peak shift, the second inequality implies that the more conspicuous a type of prey is, the more memorable it is^{18,23}. It is apparent from condition (11) that the greater the predator's memory retention ($1/c(x,u)$) and associative memory acquisition ($a(x,x,u)$) are, the more likely it is for aposematism to evolve. We note that the evolutionary dynamics in the case of increasing memory-retention toward greater conspicuousness are qualitatively the same as those in the case of peak shift (shown in Fig. 2).

be written as

$$N > p'(x)/[p(x)^2 \partial a(y,x,u)/\partial y]_{y=x} \quad (3)$$

and this relation between N and x is represented by the region above the curve line in Fig. 2 (Box 3). In this region, the conspicuousness x of the prey population should increase, whereas it should decrease outside the region (arrows in Fig. 2). Starting with the original conspicuousness value $x_0 (> 0)$ of the cryptic wild type (Box 3), conspicuousness x will evolve and lead to a higher level (indicated in Fig. 2 by the thick part of the curve line) once N exceeds a threshold value N_0 (determined by the intersection between the curve line and the line $x = x_0$ in Fig. 2). Thus, although the evolution of aposematism proceeds in a gradual manner, a great discontinuity in the degree of conspicuousness is created as a result of the evolution. This discontinuity predicted by the model can explain the unambiguous distinction in appearance between many aposematic species and their closely related cryptic species, which is observed in nature.

We have shown that a predator's generalization with peak shift, which creates a substantial associative learning of a rare mutant by the predator solely experiencing an abundant wild type, may overcome the two barriers to the initial evolution of aposematism pointed out by Guilford^{1,2}, helping the initial evolution of aposematism to occur.

It has been suggested that aposematism may evolve by gradual change²¹. Our model provides a mechanism for, and thus supports, this idea. Peak shift has been shown to be a stabilizing force once aposematism has evolved^{3,13}; our model reveals that peak shift can also be a driving force for the initial evolution of aposematism.

We have assumed that $c = 1$, but if $c = c(x,u)$, our model indicates that not only peak shift, but also increasing memory retention toward greater conspicuousness ($c_x(x,u) < 0$), can give rise to the evolution of aposematism (Box 4). □

Received 11 May; accepted 29 June 1998.

- Guilford, T. The evolution of conspicuous coloration. *Am. Nat.* **131**, S7–S21 (1988).
- Guilford, T. In *Insect Defenses* (eds Evans, D. L. & Schmidt, J. O.) 23–61 (State Univ. New York Press, 1990).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
- Harvey, P. H., Bull, J. J. & Paxton, R. J. Why some insects look pretty nasty. *New Sci.* **97**, 26–27 (1983).
- Leimar, O., Enquist, M. & Sillén-Tullberg, B. Evolutionary stability of aposematic coloration and prey unprofitability: a theoretical analysis. *Am. Nat.* **128**, 469–490 (1986).
- Mallet, J. & Singer, M. Individual selection, kin selection, and the shifting balance in the evolution of warning colours: the evidence from butterflies. *Biol. J. Linn. Soc.* **32**, 337–350 (1987).
- Gagliardo, A. & Guilford, T. Why do warning-coloured prey live gregariously? *Proc. R. Soc. Lond. B* **251**, 69–74 (1993).
- Alatalo, R. V. & Mappes, J. Tracking the evolution of warning signals. *Nature* **382**, 708–710 (1996).
- Sillén-Tullberg, B. Evolution of gregariousness in aposematic butterfly larvae: a phylogenetic analysis. *Evolution* **42**, 293–305 (1988).
- Tullberg, B. S. & Hunter, A. F. Evolution of larval gregariousness in relation to repellent defences and warning coloration in tree-feeding Macrolepidoptera: a phylogenetic analysis based on independent contrasts. *Biol. J. Linn. Soc.* **57**, 253–276 (1996).
- Sillén-Tullberg, B. The effect of biased inclusion of taxa on the correlation between discrete characters in phylogenetic trees. *Evolution* **47**, 1182–1191 (1993).

- Weary, D. M., Guilford, T. C. & Weisman, R. G. A product of discriminative learning may lead to female preferences for elaborate males. *Evolution* **47**, 333–336 (1993).
- Leimar, O. & Tuomi, J. Synergistic selection and graded traits. *Evol. Ecol.* **12**, 59–71 (1998).
- Gamberale, G. & Tullberg, B. S. Evidence for a peak-shift in predator generalization among aposematic prey. *Proc. R. Soc. Lond. B* **263**, 1329–1334 (1996).
- Hanson, H. M. Effects of discrimination training on stimulus generalization. *J. Exp. Psychol.* **58**, 321–334 (1959).
- Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
- Gittleman, J. L. & Harvey, P. H. Why are distasteful prey not cryptic? *Nature* **286**, 149–150 (1980).
- Roper, T. J. & Redston, S. Conspicuousness of distasteful prey affects the strength and durability of one-trial avoidance learning. *Anim. Behav.* **35**, 739–747 (1987).
- Owen, R. E. & Owen, A. R. G. Mathematical paradigms for mimicry: recurrent sampling. *J. Theor. Biol.* **109**, 217–247 (1984).
- Guilford, T. Evolutionary pathways of aposematism. *Oecologia* **11**, 835–841 (1990).
- Ender, J. A. Frequency-dependent predation, crypsis and aposematic coloration. *Phil. Trans. R. Soc. Lond. B* **319**, 505–523 (1988).
- Karlin, S. *A First Course in Stochastic Processes* (Academic, New York, 1969).
- Schuler, W. & Roper, T. J. Responses to warning coloration in avian predators. *Adv. Study Behav.* **21**, 111–146 (1992).

Acknowledgements. We thank N. Yamamura for discussions; O. Takeyama for technical advice; and J. Lawton and the NERC Centre for Population Biology for hospitality and discussions offered to M.H. in finalizing the manuscript. This work was supported by an MESSC grant to M.H.

Correspondence and requests for materials should be addressed to M.H. (e-mail: higashi@ecology.kyoto-u.ac.jp).

Effects of sexual dimorphism on facial attractiveness

D. I. Perrett*, K. J. Lee*†, I. Penton-Voak*, D. Rowland*, S. Yoshikawa‡§, D. M. Burt*, S. P. Henzill, D. L. Castles§ & S. Akamatsu§

* School of Psychology, University of St Andrews, Fife KY16 9JU, UK

† Department of Cognitive Psychology in Education, Graduate School of Education, Kyoto University, Sakyo, Kyoto 606-8501, Japan

§ ATR, Human Information Processing Research Laboratories, 2-2 Hikari-dai, Soraku-gun, Kyoto, 619-02, Japan

|| Department of Psychology, University of Natal, King George V Avenue, Durban 4001, South Africa

¶ Hasegawa Laboratory, Department of Life Sciences, University of Tokyo, Komaba, Tokyo 153, Japan

Testosterone-dependent secondary sexual characteristics in males may signal immunological competence¹ and are sexually selected for in several species^{2,3}. In humans, oestrogen-dependent characteristics of the female body correlate with health and reproductive fitness and are found attractive^{4–6}. Enhancing the sexual dimorphism of human faces should raise attractiveness by enhancing sex-hormone-related cues to youth and fertility in females^{5,7–11}, and to dominance and immunocompetence in males^{5,12,13}. Here we report the results of asking subjects to choose the most attractive faces from continua that enhanced or diminished differences between the average shape of female and male faces. As predicted, subjects preferred feminized to average shapes of a female face. This preference applied across UK and Japanese populations but was stronger for within-population judgements, which indicates that attractiveness cues are learned. Subjects preferred feminized to average or masculinized shapes of a male face. Enhancing masculine facial characteristics increased both perceived dominance and negative attributions (for example, coldness or dishonesty) relevant to relationships and paternal investment. These results indicate a selection pressure that limits sexual dimorphism and encourages neoteny in humans.

Computer-graphic techniques can be used to construct 'average' male and female faces by digitally blending photographs of individuals of the same sex¹⁴ (Fig. 1). Sexual dimorphism in face shape can then be enhanced or diminished^{14,15} (Fig. 2). We presented such manipulations of both Japanese and Caucasian face stimuli to

† Present address: Department of Psychology, University of Western Australia, Nedlands, Western Australia 6907, Australia.

Japanese subjects in Japan and Caucasian subjects in Scotland.

The amount of transformation (that is, masculinization or feminization) that was applied by subjects to obtain the most attractive face shape was compared with a mean of 0% predicted by the null hypothesis (that altering sex-related characteristics would not affect attractiveness) and predicted by the hypothesis that attractiveness is averageness¹⁶. The face shape selected by Caucasian subjects as most attractive (from the shape range available) was significantly feminized for both the Caucasian female face (mean level of feminization was 24.2%; $t_{49} = 7.6$, $P < 0.001$) and the Japanese female face continua (mean 10.2%; $t_{49} = 2.3$, $P = 0.027$). Japanese subjects also selected significantly feminized versions of the female stimuli for both the

Japanese (mean 22.9%; $t_{41} = 7.6$, $P = 0.001$) and the Caucasian (mean 15.3%; $t_{41} = 4.5$, $P = 0.001$) female face continua.

Three-way analysis of variance (ANOVA) of the level of transform applied by subjects to define attractive face shapes revealed no main effect of subject sex ($F_{1,88} = 1.58$, $P = 0.21$), population of subjects ($F_{1,88} = 0.32$, $P = 0.57$) or type of stimulus face (Japanese/Caucasian; $F_{1,88} = 1.42$, $P = 0.24$). The only significant interaction between main effects was that between subject population and type of stimulus face ($F_{1,88} = 17.06$, $P < 0.001$), which was attributable to the greater degree of feminization preferred for stimulus faces of the subject's own population (Fig. 3a).

Previous studies show cross-population consistency in judgements of attractiveness^{9,11,14,17}. Our study shows cross-cultural (between-population) agreement in the preference for feminized to average face shapes, which refutes the averageness hypothesis^{14,16}. The study also indicates effects of experience on judgements of female attractiveness as a greater degree of feminization was preferred for faces from the subject's own population than for faces from a different population. Both generalization and cultural specificity of judgements of attractiveness may result from learning. We find cues to female attractiveness relate to the way that female faces differ from males. Sensitivity to the consistent sex differences in faces (and hence female attractiveness) could be learned through exposure to male and female exemplars. Most differences learned this way will generalize between populations as they reflect the common action of sex hormones during growth. Subjects, however, may become more sensitive to the sexual dimorphism of faces within the subject's own population because of increased exposure to population-specific male–female variations.

For the male face stimuli, the shape selected by Caucasian subjects as most attractive (from the shape range available) was significantly feminized for both the Caucasian male face (mean level of feminization was 15%; $t_{49} = 4.22$, $P < 0.001$) and the Japanese male face continua (mean 9%; $t_{49} = 2.2$, $P = 0.03$). Japanese subjects also selected significantly feminized versions of the male stimuli for both the Japanese (mean 20%; $t_{41} = 6.5$, $P < 0.001$) and the Caucasian (mean 17%; $t_{41} = 4.8$, $P < 0.001$) male face continua. For the male stimuli, three-way ANOVA revealed there was no main effect of subject sex ($F_{1,88} = 0.18$, $P = 0.67$), subject population ($F_{1,88} = 2.94$, $P = 0.09$) or type of stimulus face ($F_{1,88} = 0.02$, $P < 0.89$) and no significant interactions between effects.

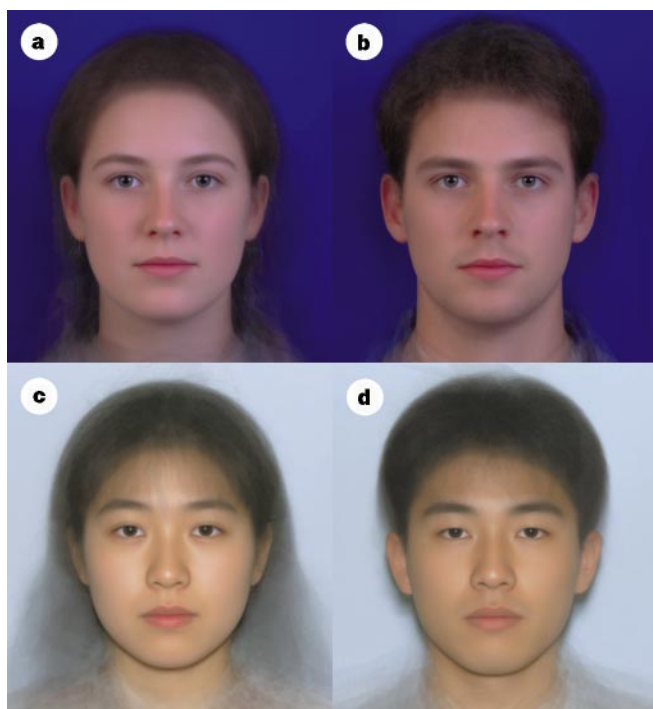


Figure 1 Composite 'average' facial images. **a**, 'Caucasian' female face; **b**, 'Caucasian' male face; **c**, 'Japanese' female face; **d**, 'Japanese' male face.

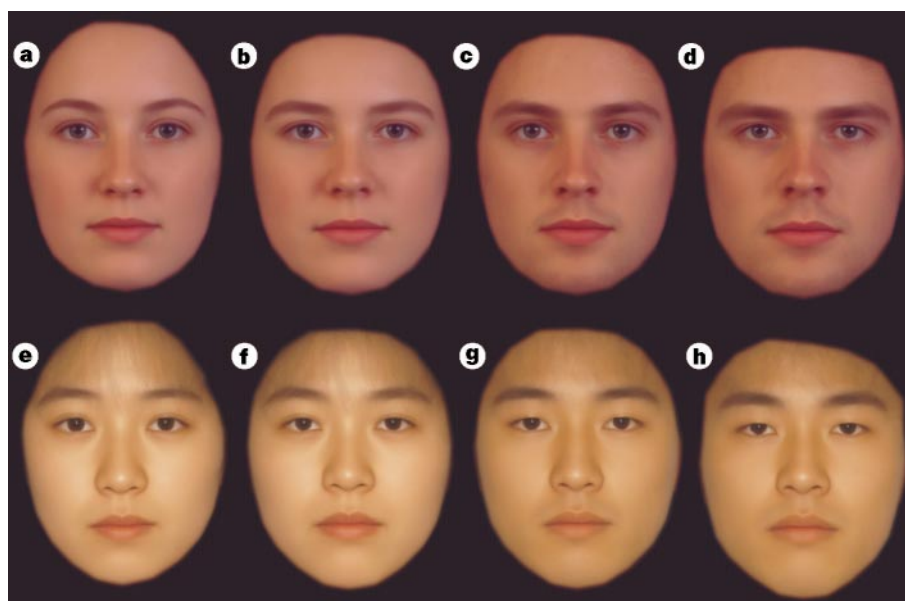


Figure 2 Facial images of Caucasian and Japanese females and males that were 'feminized' and 'masculinized' 50% in shape. **a**, Caucasian female, feminized; **b**, Caucasian female, masculinized. **c**, Caucasian male, feminized; **d**, Caucasian

male, masculinized. **e**, Japanese female, feminized; **f**, Japanese female, masculinized. **g**, Japanese male, feminized; **h**, Japanese male, masculinized.

Asymmetries in the facial outline (from the hairline), which remain after cropping, could contribute to judgements. With a different set of Caucasian faces (19 male, 17 female, 30–35 years old), symmetrical composites were made by averaging component faces and their mirror reflections. Caucasian subjects ($n = 67$, age range 15–40, 23 female) made forced-choice judgements of attractiveness of symmetrical average stimuli that were 50% masculinized or feminized. Masculinization of face shape decreased attractiveness of male (87% of subjects; Binomial test $P = 0.001$) and female faces (78%; $P < 0.001$), whereas feminization increased attractiveness of male (64%; $P < 0.05$) and female faces (53%; $P < 0.05$).

Males have larger faces than females. However, standardizing the distance between pupils removes this size difference. We prepared composite images from a new set of Caucasian faces (26 male, 17 female, 18–21 years old) without standardizing the inter-pupil separation. Manipulation of these composites maintained sexual dimorphism in face shape and size. Caucasian subjects ($n = 135$, age range 15–71, 65 female) ranked average images that were masculinized and feminized by 50% for attractiveness. Masculinization of the average shape decreased attractiveness of male (74% of subjects; $P < 0.000005$) and female (76%; $P < 0.000005$) faces, whereas feminization increased attractiveness rankings for male (58%; $P = 0.029$) and female (60%; $P < 0.013$) faces.

Thus, preference for feminized face shapes over average male and female face shapes was found with interactive and forced-choice methods using different face sets, even when the potential contributions by symmetry and size dimorphism were controlled.

To interpret preferences, 50% masculinized, 50% feminized and cropped average images (Fig. 2) were rated for perceived characteristics by a new set of subjects. Twenty Caucasian subjects (age range

18–50, 10 female) were presented with four sets of three images that represented the end points of each continuum and the average. Subjects were asked to rank stimuli from one set on seven characteristics (masculinity, dominance, warmth, emotionality, honesty, intelligence and age). The order of testing of characteristics and image sets was randomized. An additional 20 subjects (age range 19–61, 10 female) ranked the stimuli on three further characteristics (cooperativeness, assertiveness and 'good parent').

For Caucasian and Japanese male faces, increasing the masculinity of face shape across the three set members increased ranking of perceived dominance, masculinity and age but decreased ranking of perceived warmth, emotionality, honesty, cooperativeness and quality as a parent (Friedman's $\chi^2 \geq 15.6$, degrees of freedom (d.f.) = 2, $P < 0.0005$, for each rated dimension). Increasing masculinity affected the Japanese and Caucasian female face sets in the same way for all characteristics ($\chi^2 \geq 8.1$, d.f. = 2, $P < 0.017$, each dimension), except for 'good parent' with the Caucasian female faces, where the rank order was average, feminized and masculinized ($\chi^2 \geq 6.7$, d.f. = 2, $P < 0.035$). Increasing masculinity did not consistently decrease apparent intelligence (Caucasian male and female faces, $P > 0.5$; Japanese female face, $P = 0.07$; Japanese male face, $P = 0.02$) or increase attributions of assertiveness (Japanese and Caucasian female faces, $P > 0.5$; Japanese male face, $P = 0.058$; Caucasian male face, $P = 0.157$).

The preference for male face shapes that are slightly feminized may reflect the effects of masculinity on perceived age. Whereas status and height are valued in males^{9,18}, youth benefits judgements of attractiveness for both female^{9,11,19} and male¹⁹ faces. For both males and females, enhancing sexual dimorphism in face shape develops cues to characteristics which, from a biological perspective, appear beneficial (that is, youth and fertility in females^{7–11} and dominance in males^{12,13,18}). For males, however, enhancing masculinity in face shape also predisposes some negative personality attributions. Such attributions, although stereotypic, may predict behaviour; ratings of perceived dishonesty from facial appearance correlate with the face owner's willingness to participate in deceptive behaviour²⁰. Indeed, increasing testosterone level in males is associated with more troubled relationships (including increased infidelity, violence and divorce)²¹. Feminization of male face shape may increase attractiveness because it 'softens' particular features^{10,22} that are perceived to be associated with negative personality traits.

Together, the results indicate that judgements of male attractiveness reflect multiple motives²². Females may adopt different strategies, giving preference to characteristics that are associated with dominance and an effective immune system^{12,13}, or to characteristics that are related to paternal investment.

Sexual dimorphism in any species reflects compromises among diverse selection pressures. In humans, the greater upper body musculature and more rugged skeletal anatomy of males relative to females may reflect advantages in male–male competition and hunting. Because male attractiveness is an important determinant of relationships and sexual partnerships²³, the reduction in attractiveness of male face shape with masculinization represents a further selection pressure. This would act against 'run away' fisherian sexual selection for extreme male characteristics¹, and is consistent with the relative lack of sexual dimorphism in humans²⁴.

The preferences found here indicate a selection pressure on the evolution of face shape that acts against pronounced differences between males and females and, as more-feminine face shapes are perceived as younger, the preferences would encourage a youthful, neotinous appearance in the species generally. □

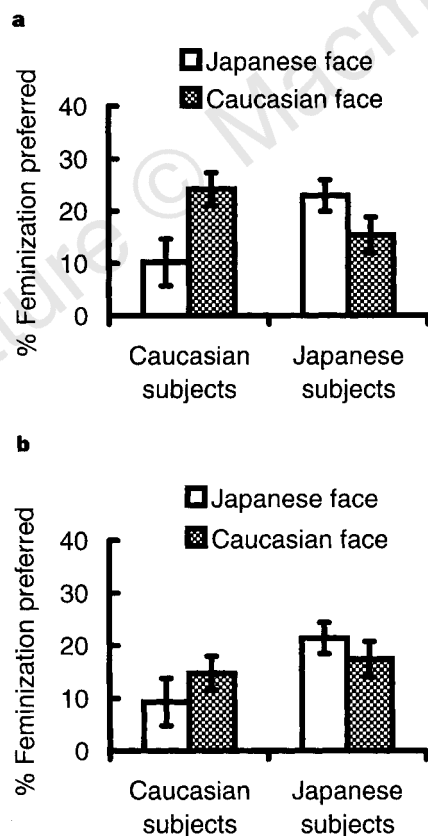


Figure 3 The effect of feminization of face shape on judgements of female and male attractiveness. **a**, Female stimuli; **b**, male stimuli. Overall, subjects preferred a feminine face shape to an average shape both within and between populations. The degree of feminization preferred was greater within than between populations for female faces.

Methods

Preparation of composite facial images. Japanese faces (students at Ottemon-Gakuin University; 28 male, age 20–23 years, mean 21.6 years; 28 female, age 20–22 years, mean 21.4 years) were photographed under standard lighting conditions with neutral facial expression. Similar photographs were

prepared for Caucasian faces (students at St Andrews University, 25 male, age 19–23 years, mean 21.0 years; 30 female, age 19–22 years, mean 20.6 years). Photographs were converted to digital format (Kodak Photo-CD) and 174 feature points on salient facial landmarks (for example, nose-tip) were defined manually for each face^{14,15}. The average face shapes of the male and female Japanese and Caucasian face subsets were calculated from the feature points. The position of eye centres was standardized for corresponding average male and female face shapes. Each original face image was then warped to the shape of the corresponding average face and the resultant reshaped face images were blended together by averaging colour and intensity values of pixels at corresponding image locations^{14,15} (Fig. 1). The vector difference between corresponding feature points on the male and female averages was increased or decreased by 50% to create feminized and masculinized shapes. The image of the composite face was then warped into these new face shapes to create image pairs with identical texture but enhanced or diminished sexually dimorphic differences in face shape. The size of all male and female face images was matched by standardization of inter-pupil distance. The resulting composite images were cropped around the face and faded into a black background (Fig. 2). Cropping removed the hair, ears and neck, which were not consistent in shape or visibility in component images because of differing hairstyles and clothing.

Procedure. A Silicon Graphics Indigo² Maximum Impact (4 MB TRAM) was used to create smooth continua between 50% masculinized and 50% feminized face pairs (Fig. 2) as the end points, and the cropped average as the midpoint. The point along a shape continuum was controlled interactively by the position of the computer mouse. The appropriate image was calculated in real-time using texture mapping hardware. Stimuli were presented in 24-bit colour at the centre of an 800 × 800 pixel window. Fifty Caucasian subjects (research staff and students from St Andrews University; age 19–31 years, 25 female) and 42 Japanese subjects (research staff and students from ATR and Doshisha University; age 18–44 years, 19 female) were instructed to select the most attractive face from the continuum. Each continuum was presented twice to allow left/right counterbalancing of the end points, making a total of eight trials in randomized order.

Received 16 January; accepted 10 July 1998.

- Kirkpatrick, M. & Ryan, M. J. The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33–38 (1991).
- Andersson, M. Female choice for extreme tail length in a widowbird. *Nature* **299**, 818–820 (1992).
- Møller, A. P. Female swallow preference for symmetrical male sexual ornaments. *Nature* **357**, 238–240 (1992).
- Singh, D. Body shape and women's attractiveness—the critical role of waist-to-hip ratio. *Hum. Nature* **4**, 297–3221 (1993).
- Barber, N. The evolutionary psychology of physical attractiveness: sexual selection and human morphology. *Ethol. Sociobiol.* **16**, 395–424 (1995).
- Manning, J. T., Scutt, D., Whitehouse, G. H. & Leinster, S. J. Breast asymmetry and phenotypic quality in women. *Evol. Hum. Behav.* **18**, 2223–236 (1997).
- Symons, D. *The Evolution of Human Sexuality* (Oxford Univ. Press, 1979).
- Cunningham, M. R., Measuring the physical in physical attractiveness: quasi-experiments on the sociobiology of female facial beauty. *J. Pers. Soc. Psychol.* **50**, 925–935 (1986).
- Buss, D. M. Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behav. Brain Sci.* **122**, 1–49 (1989).
- Johnston, V. S. & Franklin, M. Is beauty in the eye of the beholder? *Ethol. Sociobiol.* **14**, 183–199 (1993).
- Jones, D. Sexual selection, physical attractiveness, and facial neoteny. *Curr. Anthropol.* **36**, 723–748 (1995).
- Grammer, K. & Thornhill, R. Human (*Homo sapiens*) facial attractiveness and sexual selection: the role of symmetry and averageness. *J. Comp. Psychol.* **108**, 233–242 (1994).
- Thornhill, R. & Gangestad, S. The evolution of human sexuality. *Trends Ecol. Evol.* **11**, 98–102 (1996).
- Perrett, D. I., May, K. A. & Yoshikawa, S. Facial shape and judgements of female attractiveness. *Nature* **368**, 239–242 (1994).
- Rowland, D. A. & Perrett, D. I. Manipulating facial appearance through shape and color. *IEEE Comput. Graph. Appl.* **15**, 70–76 (1995).
- Langlois, J. H. & Roggman, L. A. Attractive faces are only average. *Psychol. Sci.* **1**, 115–121 (1990).
- Cunningham, M. R., Roberts, A. R., Barbee, A. P. & Druen, P. B. "Their ideas of beauty are, on the whole, the same as ours": consistency and variability in the cross-cultural perception of female attractiveness. *J. Pers. Soc. Psychol.* **68**, 261–279 (1995).
- Jackson, L. A. *Physical Appearance and Gender: Sociobiology and Sociocultural Perspectives* (State Univ. New York Press, Albany, 1992).
- Deutsch, F. M., Zalsenski, C. M. & Clark, M. E. Is there a double standard of ageing. *J. Appl. Soc. Psychol.* **16**, 771–785 (1986).
- Berry, D. S. & Wero, J. L. F. Accuracy of face perception: a view from ecological psychology. *J. Pers.* **61**, 497–423 (1993).
- Booth, A. & Dabbs, J. Testosterone and men's marriages. *Social Forces* **72**, 463–477 (1993).
- Cunningham, M. R., Barbee, A. P. & Pike, C. L. What do women want? Facialmetric assessment of multiple motives in the perception of male facial attractiveness. *J. Pers. Soc. Psychol.* **59**, 61–72 (1990).
- Gangestad, S. W. & Thornhill, R. The evolutionary psychology of extrapair sex: the role of fluctuating symmetry. *Ethol. Hum. Behav.* **18**, 69–88 (1997).
- Martin, R. D. & May, R. M. Outward signs of breeding. *Nature* **293**, 7–9 (1990).

Acknowledgements. This work was supported by Unilever Research and the ESRC. We thank A. Whiten, R. Byrne, R. Barton, J. Lycett, S. Reicher, D. Carey, M. Ridley, J. Graves and D. Symons for comments.

Correspondence and requests for materials should be addressed to D.I.P. (e-mail: dp@st-andrews.ac.uk).

Separate body- and world-referenced representations of visual space in parietal cortex

Lawrence H. Snyder*, Kenneth L. Grieve*, Peter Brothchie* & Richard A. Andersen

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

In order to direct a movement towards a visual stimulus, visual spatial information must be combined with postural information¹. For example, directing gaze (eye plus head) towards a visible target requires the combination of retinal image location with eye and head position to determine the location of the target relative to the body. Similarly, world-referenced postural information is required to determine where something lies in the world. Posterior parietal neurons recorded in monkeys combine

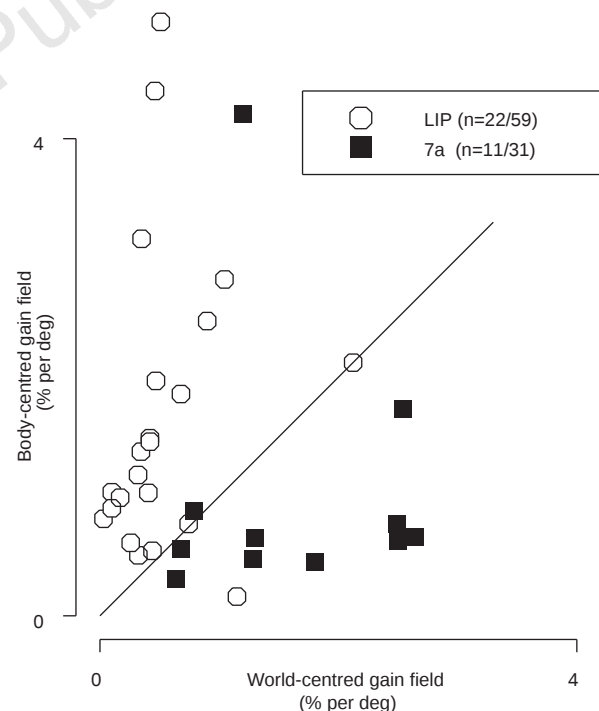


Figure 1 LIP responses (open circles) were modulated by body- but not world-referenced target location; 7a responses (filled squares) were modulated by world- but not body-referenced location. Visually evoked or delayed saccades were made after combined head-and-body rotation in the dark (world-referenced modulation), or after an equal counter-rotation of the body under a stable head. The absolute value of the gain field is shown for cells whose responses during and immediately after visual cue presentation depended on either body- or world-referenced head position (Student's *t*-test, $P < 0.05$; 33 of 90 cells). Cells above the diagonal line had stronger body-referenced modulation; cells below the line had stronger world-referenced modulation. Only two 7a cells fell above the line, and only two LIP cells fell below the line, indicating that body- and world-referenced modulation were well segregated by area (see Table 1).

* Present addresses: Department of Anatomy and Neurobiology, Box 8108, Washington University School of Medicine, 660 S Euclid Avenue, St Louis, Missouri 63110, USA (L.H.S.); Department of Optometry and Vision Sciences, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK (K.L.G.); Medical Imaging Department, St Vincent's Hospital, Princess Street, Fitzroy, Victoria 3065, Australia (P.B.).

visual information with eye and head position²⁻⁴. A population of such cells could make up a distributed representation of target location in an extraretinal frame of reference⁴⁻⁷. However, previous studies have not distinguished between world-referenced and body-referenced signals^{4,8}. Here we report that modulations of visual signals (gain fields) in two adjacent cortical fields, LIP and 7a, are referenced to the body and to the world, respectively. This segregation of spatial information is consistent with a streaming of information, with one path carrying body-referenced information for the control of gaze, and the other carrying world-referenced information for navigation and other tasks that require an absolute frame of reference.

Head position can be referenced to either the body or to the world. A body-referenced frame could derive from neck proprioception, or from efference copy of neck-muscle motor commands. World-referenced gain fields could derive from vestibular or visual signals. Parietal cells might use just one of these frames of reference; both frames could coexist on the same population of cells; or the two frames could be coded by two distinct cell populations⁹⁻¹².

We devised three tests to distinguish between these possibilities in rhesus macaques. To isolate body-referenced gain fields, visual responses to targets presented at identical retinal locations were compared after the body had been counter-rotated underneath the head to one of 2-7 positions (body-under-head rotation); after counter-rotation, the head was always at the same angle with respect to the world, but at different angles with respect to the body. To isolate world-referenced gain fields, responses to retinally identical targets were compared after the head and body had been rotated together to one of the same 2-7 positions (body-plus-head rotation);

after combined rotation, the head was always at the same angle (0°) with respect to the body, but was at varying angles with respect to the world. Finally, for active head rotation, responses to retinally identical targets were compared after the animal had oriented its head to one of 2-5 positions; after active head rotation, head position varied with respect to both the world and the body.

We recorded from 288 cells in two macaques. Every trial began with an interval of central fixation followed by a peripheral target, always at the same retinotopic position. One animal made an immediate saccade to the target (231 cells), and the second memorized the target's location and acquired it after a short delay (57 cells). Responses to retinally identical targets were compared on interleaved trials following different head and body movements.

First we investigated whether head position was encoded with respect to the body or world. Ninety cells were tested using interleaved body-under-head and body-plus-head rotations. Many cells showed either body- or world-referenced gain fields, but few cells showed both. In addition, cells with body- and world-referenced gain fields were anatomically segregated. In cortical area LIP, 22 of 59 cells showed significant ($P < 0.05$) modulation for at least one of the two rotations: 20 body-referenced, one world-referenced, and only one mixed body- and world-referenced. In area 7a, 11 of 31 cells showed significant modulation: nine world-referenced, one body-referenced, and one mixed.

The segregation of body-referenced gain fields in LIP and world-referenced gain fields in 7a is shown in Fig. 1. Each point represents a cell with significant modulation in at least one rotation, coded by area. Gain fields were quantified by the mean percentage change in response to a 1° change in body-under-head (ordinate) or body-

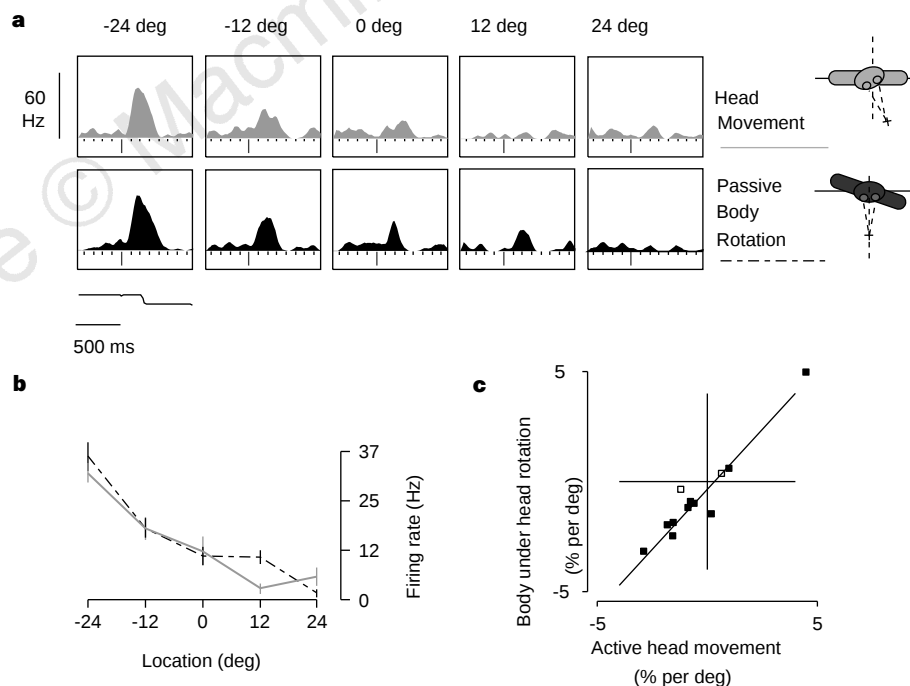


Figure 2 Head-position gain fields in LIP were accounted for by body-referenced modulation. **a**, On interleaved trials, animals made identical visual saccades from five different initial fixation points using two different head and body postures. Half of the saccades occurred with the body straight ahead and the head turned to one side (top row), the other half with the head straight ahead and the body turned to one side (bottom row). For each inset, data from eight saccades were smoothed and aligned on target presentation (large tick). An example eye-position trace is shown at the bottom left. **b**, The modulation resulting from changing both head-on-body and head-in-world position (solid grey line) was nearly identical to the gain field resulting from changing only head-on-body position (broken black line). Gain-field strengths (4.0 and 4.2%/per deg, respectively)

were calculated by fitting a least-squares regression to the mean discharge 50-450 ms after target appearance. Error bars show ± 1 s.e.m. **c**, Similar results were obtained in the 10 out of 27 LIP cells tested with significant body-referenced gain fields (filled squares); the gain field obtained by varying head position only with respect to the body and not with respect to the world (ordinate) equalled that produced by changing both body- and world-referenced head position (abscissa). The line shows the result of a least-squares linear regression of the data (1.06 ± 0.10 , not significantly different from 1 [$P = 0.56$]; $r^2 = 0.92$). Two additional cells had significant gain-field modulation for the active head but not the body-under-head rotation (open squares). Least-squares regression through all 12 cells yielded a regression coefficient of 1.09 ± 0.09 ($P = 0.34$; $r^2 = 0.52$).

plus-head position (abscissa), performed in the dark. Most 7a cells fell beneath the diagonal, indicating that world-referenced gain fields were systematically stronger than body-referenced gain fields. Most LIP cells fell on or above the diagonal, indicating that body-referenced gain fields were systematically greater than or equal to world-referenced fields. Out of 31 7a and 59 LIP cells tested, only two cells showed significant effects in both reference frames. In a separate series of experiments in which only head-plus-body rotation was performed (see Table 1), world-referenced modulation was more than twice as frequent in 7a than in LIP (40% versus 18% of cells, respectively), and the mean absolute modulation in cells with significant effects was approximately 50% greater in 7a than in LIP (1.71 ± 0.17 versus $1.11 \pm 0.13\%$ modulation per degree of head rotation, mean $\pm$ 1 s.e.m.).

Thus, when the head was repositioned with respect to the body, visual responses in LIP were modulated. When the head was repositioned with respect to the world, visual responses in 7a were modulated. However, active head movements typically displace the head in both frames of reference. The next two experiments examined gain-field modulation when the head was actively rotated with respect to both the body and the head.

In LIP, body-referenced gain-field effects could entirely account for the modulation seen after active head rotations (Fig. 2). Gain fields were compared after body-under-head and active head rotations. The top row of peristimulus time histograms in Fig. 2a shows visual responses to targets presented after an active head rotation to one of five positions. With the head turned to the left, target appearance elicited a robust response. With the head turned to the right, a retinotopically identical stimulus elicited no response. The bottom row shows that gain fields obtained after counter-rotating the body under the head were virtually identical (Fig. 2b), indicating that the gain field seen for this neuron after active head movement could be explained completely by body-referenced head position information, with little or no contribution from vestibular signals. Equal contributions from efference copy signals under both conditions cannot be ruled out, because it is possible (although unlikely) that the animal actively assisted during the body-under-head movement.

Most LIP cells had body-referenced gain fields (Table 1). Gain fields measured after active head and body-under-head rotations are compared in Fig. 2c. Of 27 cells, 12 showed a significant effect for at least one rotation. A regression analysis failed to show a difference between the two conditions, showing that LIP gain fields were

Table 1 Cells in LIP and 7a with significant body- or world-referenced gain fields

Area	Task	Total	Body	World	Both (mixed)
LIP	Memory saccades (world plus body)	32	10	0	1
	Visual saccades (world plus body)	27	10	1	0
	Visual saccades (world only)	50		9	
	Total		20 of 59 (34%)	10 of 109 (9%)	1 of 59 (2%)
7a	Memory saccades (world plus body)	25	0	7	0
	Visual saccades (world plus body)	6	1	2	1
	Visual saccades (world only)	148		60	
	Total		1 of 31 (3%)	69 of 179 (39%)	1 of 31 (3%)

In LIP, most head-position gain fields were body referenced; in 7a, most were world referenced. Few cells showed mixed effects. The use of memory or visual saccades yielded similar results. Cells were tested for either both world- and body-referenced effects (world plus body) or for world-referenced effects alone (world only). The distribution of cells with significant effects as a function of anatomical area was statistically significant (chi-squared test on body-referenced modulation: $\chi^2 = 9.09$, 1; $P = 0.0026$; chi-squared test on world-referenced modulation, with Yates' correction: $\chi^2 = 12.271$, 1; $P = 0.00046$).

driven primarily by body-referenced signals, with little or no contribution from world-referenced vestibular signals, even during active head movement (least-squares regression coefficient: 1.09 ± 0.09 ; $r^2 = 0.95$).

In a third experiment, we investigated whether world-referenced gain fields in 7a could entirely account for the modulation seen after active rotations of the head relative to the body. Although we found no effect of passive rotation of the body under the head, it is possible that proprioceptive or efference copy signals might nonetheless influence 7a cells when delivered in conjunction with a vestibular signal. Gain fields obtained after active head rotation were compared with those obtained after body-plus-head rotation. World-referenced signals account for only part of the effect during active head rotation (Fig. 3). Of 102 cells tested, 57 showed significant gain fields in at least one condition. Regression analysis revealed a coefficient relating the two gain fields of 0.61 ± 0.10 , with a large amount of residual variance ($r^2 = 0.40$). Of these 57 cells, 18 showed significant modulation after active head but not after passive body-plus-head rotation, suggesting an influence of proprioception or efference copy. Alternatively, the faster speed of head rotation in the active head rotation task might have produced a more robust vestibular signal.

To confirm the vestibular origin of world-centred gain-field modulations in 7a, we compared responses after small rotations (± 6 – 12°), either above (10° per s^2) or below (0.04 – 0.08° per s^2) the vestibular threshold. In six of seven cells, gain fields appeared after rotation above but not below threshold, strongly suggesting a vestibular input. In the absence of direct connections with the brainstem vestibular nuclei, the origin of this input could be either cortical or subcortical from vestibular-related areas. Because the vestibular canals signal angular velocity, whereas gain fields depend

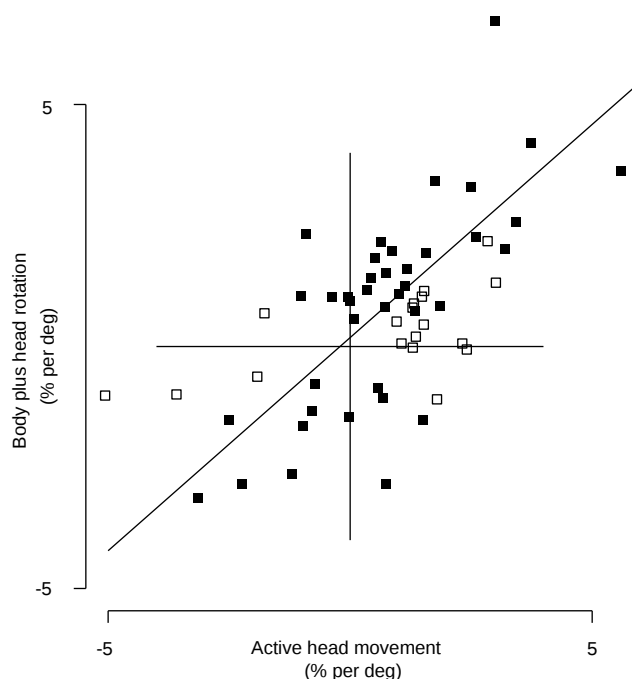


Figure 3 The 7a world-referenced gain field modulation did not fully account for modulation seen after active head movements. Gain-field modulation strengths are compared for body-plus-head (ordinate) and active head rotation (abscissa). Of 102 cells, 39 showed significant world-referenced gain-field modulation after body-plus-head rotation (filled squares). The solid diagonal shows least-squares linear regression through these data (0.88 ± 0.14 , not significantly different from 1 [$P = 0.39$]; $r^2 = 0.52$). An additional 18 cells had significant gain-field modulation for the active head but not the body-plus-head rotation (open squares). Least-squares regression through all 57 cells yielded a coefficient of 0.61 ± 0.10 ($P < 0.01$; $r^2 = 0.40$).

on position, a mathematical integration of the afferent vestibular signal would be required¹³. Note that the efference copy of a velocity command would require a similar integration¹⁴.

LIP is more closely associated than 7a with saccadic eye movements. LIP projects more heavily upon the frontal eye field than does 7a, and LIP responses are more often presaccadic than 7a responses^{15,16}. Electrical microstimulation of LIP elicits saccadic eye movements, whereas stimulation of 7a does not^{17,18}. LIP, but not 7a, has been shown to play a specific role in saccadic eye-movement intention¹⁹. The finding of body-referenced modulation in LIP is surprising, because retinotopic (eye-referenced) target location would suffice for planning eye movements. However, emerging evidence suggests that the apparent emphasis on the coding of eye movements in isolation may be an artefact of using a head-fixed preparation. Primates normally move their head and even their trunk to shift their gaze, so gaze centres in the primate may control more than just the eye²⁰. Our finding of body-referenced modulation of visual signals in LIP is consistent with a more general role in directing gaze to visible and remembered targets.

Area 7a projects heavily to the parahippocampal gyrus and the presubiculum, regions known in primates to be involved in the generation of topographical memory, and in rats to be involved in world-referenced navigation^{21–24}. The finding of world-referenced gain fields in 7a is consistent with a role in such functions. Some cells in primate hippocampus are activated when the animal gazes in a particular world-referenced direction; such a pattern of activation could result from world-referenced gaze direction provided by area 7a²⁵.

A population code can store information from several frames of reference^{4,7}, yet we have shown that body- and world-referenced information is largely segregated in parietal cortex. A similar anatomical segregation occurs in the coding of spatial information for directing eye versus arm movements¹⁹. Although it is clear that anatomical segregation based on function occurs in motor areas, it has been argued that segregation in striate and extrastriate cortex is based instead on the sensory properties being analysed²⁶, or on very general functional grounds²⁷. We propose that, in posterior parietal cortex, information is segregated based on the specific functional role it will serve. The segregation of neck proprioceptive and vestibular information related to head movement is particularly noteworthy given that these two signals appear to be combined as early as the vestibular nuclei²⁸. Segregation in the cortex may reflect different computational constraints on information processing, and may thereby increase computational efficiency and help avoid the 'curse of dimensionality' that results when attempting to code many variables within a single group of cells. □

Methods

Rotations. Data were recorded from four hemispheres of two adult male rhesus macaque monkeys. A computer-controlled, motorized machinist's turntable was used to align the turntable in the world, to which the animal's body was loosely coupled. (Both animals were under constant infrared video observation to check that the animal did not twist excessively relative to the turntable, and the second animal was equipped with a shoulder harness that completely eliminated such movement.) A pair of computer-controlled brakes acting on the head-holder allowed head fixation relative to either body or world. With neither brake engaged, the head was free to rotate ± 45 degrees in the yaw plane, and to move up and down ± 0.5 cm. Animals viewed a panel of light-emitting diodes, and were trained to orient their heads towards a blinking target and to make eye movements to the locations of visible and remembered steady targets. In some experiments (such as the memory saccade data of Fig. 1), all body and head movements were passive, and different postures were achieved by sequences of turntable rotations with appropriate braking. In other experiments (such as those of Figs 2 and 3), active head movements were used instead of the brake system. In this case, passive body rotations were followed by active orienting head movements. Rotations typically required from 1.5 to 3 s to complete. The positions of the eyes, head and body were electronically monitored and recorded.

Neuronal recording. Preferred vectors of isolated cells were selected from 8 possible saccade directions and 1–2 amplitudes. The preferred vector was then used to test visual saccade responses carried out at 3–5 different gaze positions, or memory saccade responses carried out at 2 different gaze positions. Different gaze positions were established by presenting a fixation target aligned with the head after positioning the head and body in one of three postures: body straight and head deviated eccentrically; body deviated eccentrically and head aligned with body; or head straight and body deviated. Typically, 8–12 trials were performed using each gaze position from each of 2 or 3 different postures. In 14 cells, individual trials were completely randomly interleaved; in 5 cells, different postures were tested in separate blocks of trials; and in the remaining cells, 2–4 target presentations were performed after each (randomly interleaved) change in posture. These differences did not systematically affect our results. For every cell, equal numbers of oppositely directed saccades were interleaved to ensure attention to retinotopic target location. Body rotations were performed in the dark or light; saccade performance and data collection occurred only in darkness.

Data analysis. Gain fields were calculated by fitting a least-squares regression to spike rate. In memory trials, rate was measured during and immediately after visual stimulus presentation (50–350 ms after the onset of a 100-ms cue). In visual saccade trials, rate was measured in a computer-selected 400-ms interval referenced between 250 and 600 ms after cue presentation; the same interval was used for all trial types from a single cell. Similar results were obtained using a fixed, presaccadic interval. LIP and 7a were distinguished by position and depth of the recording site, and by the character of cellular responses¹⁶. Assignments to either LIP or 7a were always made before testing for the coordinate frame, and all analyses were performed in an area-blind manner.

Received 20 April; accepted 9 June 1998.

- Andersen, R. A., Snyder, L. H., Li, C. S. & Stricanne, B. Coordinate transformations in the representation of spatial information. *Curr. Opin. Neurobiol.* **3**, 171–176 (1993).
- Andersen, R. A. & Mountcastle, V. B. The influence of the angle of gaze upon the excitability of the light sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **3**, 532–548 (1983).
- Anderson, R. A., Essick, G. K. & Siegel, R. M. Encoding of spatial location by posterior parietal neurons. *Science* **230**, 456–458 (1985).
- Brochier, P. R., Andersen, R. A., Snyder, L. H. & Goodman, S. J. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* **385**, 232–235 (1995).
- Zipser, D. & Andersen, R. A. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* **331**, 679–684 (1988).
- Salinas, E. & Abbott, L. F. Transfer of coded information from sensory to motor networks. *J. Neurosci.* **15**, 6461–6474 (1995).
- Pouget, A. & Sejnowski, T. J. Spatial representations in the parietal cortex may use basis functions. *Adv. Neural Inf. Process.* **7**, 157–164 (1995).
- Duhamel, J.-R., Bremmer, F., BenHamed, S. & Graf, W. Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature* **389**, 845–848 (1997).
- Von Holst, E. & Mittelstaedt, H. Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* **37**, 464–475 (1950). (In German.)
- Mergner, T., Nardi, G. L., Becker, W. & Decelle, L. The role of canal-neck interaction for the perception of horizontal trunk and head rotation. *Exp. Brain Res.* **49**, 198–208 (1983).
- Roll, R., Velay, J. L. & Roll, J. P. Eye and neck proprioceptive messages contribute to the spatial coding of retinal input in visually oriented activities. *Exp. Brain Res.* **85**, 423–431 (1991).
- Karnath, H.-O., Sievering, D. & Fetter, M. The interactive contribution of neck muscle proprioception and vestibular stimulation to subjective "straight ahead" orientation in man. *Exp. Brain Res.* **101**, 140–146 (1994).
- Jones, G. M. & Milsum, J. H. Spatial and dynamic aspects of visual fixation. *IEEE Trans. Biomed. Eng.* **12**(2), 54–62 (1965).
- Robinson, D. A. In *Basic Mechanisms of Ocular Motility and their Clinical Implications* (eds Lennestrand, G. & Bach-y-Rita, P.) 337–374 (Pergamon, Oxford, 1975).
- Andersen, R. A., Asanuma, C., Essick, G. & Siegel, R. M. Corticocortical connections of anatomically and physiologically defined sub-divisions within the inferior parietal lobule. *J. Comp. Neurol.* **296**, 65–113 (1990).
- Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. Saccade-related activity in the lateral intraparietal area. I. Temporal properties. *J. Neurophysiol.* **66**, 1095–1108 (1991).
- Shibutani, H., Sakata, H. & Hyvarinen, J. Saccade and blinking evoked by microstimulation of the posterior parietal association cortex of monkey. *Exp. Brain Res.* **55**, 1–8 (1984).
- Thier, P. & Andersen, R. A. Electrical microstimulation suggests two different forms of representation of head-centered space in the intraparietal sulcus of rhesus monkeys. *Proc. Natl Acad. Sci. USA* **93**, 4962–4967 (1996).
- Snyder, L. H., Batista, A. P. & Andersen, R. A. Coding of intention in the posterior parietal cortex. *Nature* **386**, 167–170 (1997).
- Freedman, E. G. & Sparks, D. L. Activity of cells in the deeper layers of the superior colliculus of the rhesus monkey: evidence for a gaze displacement command. *J. Neurophysiol.* **78**, 1669–1690 (1997).
- Aguirre, G. K., Detre, J. A., Alsop, D. C. & D'Esposito, M. The parahippocampus subserves topographic learning in man. *Cereb. Cortex* **6**, 823–829 (1996).
- Squire, L. R. Memory and the hippocampus: a synthesis of findings from rat, monkey and humans. *Psychol. Rev.* **99**, 195–231 (1992).
- McNaughton, B. L., Barnes, C. A. & O'Keefe, J. The contributions of position, direction and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp. Brain Res.* **52**, 41–49 (1983).
- Baylis, G. C. & Moore, B. O. Hippocampal lesions impair spatial response selection in the primate. *Exp. Brain Res.* **98**, 110–118 (1994).
- Rolls, E. T. & O'Mara, S. M. View-responsive neurons in the primate hippocampal complex. *Hippocampus* **5**, 409–424 (1995).
- Van Essen, D. C. & Maunsell, J. H. R. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci.* **6**, 370–375 (1983).

27. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25 (1992).
28. Fredrickson, J. M., Schwarz, D. & Kornhuber, H. H. Convergence and interaction of vestibular and deep somatic afferents upon neurons in the vestibular nuclei of cat. *Acta Otolaryngol.* 61, 168–188 (1966).

Acknowledgements. We thank G. Roberts and B. Gillikin for technical assistance, and S. Gertmanian for editorial assistance. This work was supported by the Della Martin, McDonnell-Pew and Sloan foundations, the National Eye Institute and the Office of Naval Research.

Correspondence and requests for materials should be addressed to R.A.A. (e-mail: andersen@vis.caltech.edu).

Spatial exploration induces a persistent reversal of long-term potentiation in rat hippocampus

Lin Xu^{†,†}, Roger Anwyl[‡] & Michael J. Rowan^{*}

^{*}Department of Pharmacology and Therapeutics and [‡]Department of Physiology, Trinity College, Dublin 2, Ireland

[†]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, The People's Republic of China

Experience-dependent long-lasting increases in excitatory synaptic transmission in the hippocampus are believed to underlie certain types of memory^{1–3}. Whereas stimulation of hippocampal pathways in freely moving rats can readily elicit a long-term potentiation (LTP) of transmission that may last for weeks, previous studies have failed to detect persistent increases in synaptic efficacy after hippocampus-mediated learning^{4–6}. As changes in synaptic efficacy are contingent on the history of plasticity at the synapses⁷, we have examined the effect of experience-dependent hippocampal activation on transmission after the induction of LTP. We show that exploration of a new, non-stressful environment rapidly induces a complete and persistent reversal of the expression of high-frequency stimulation-induced early-phase LTP in the CA1 area of the hippocampus, without affecting baseline transmission in a control pathway. LTP expression is not affected by exploration of familiar environments. We found that spatial exploration affected LTP within a defined time window because neither the induction of LTP nor the maintenance of long-established LTP was blocked. The discovery of a novelty-induced reversal of LTP expression provides strong evidence that extensive long-lasting decreases in synaptic efficacy may act in tandem with enhancements at selected synapses to allow the detection and storage of new information by the hippocampus.

To study the effects of processing new information on the persistence of LTP in the hippocampus, we chose a task that is known to involve activation of this brain region, exploration of a new environment^{8,9}. Familiar and novel environments consisted of two boxes that were clearly distinguishable on the basis of lighting (familiar, bright versus novel, dim; see Methods). We chose to use the darker box as the novel environment because of the well known preference of rats for dimly lit areas, thereby increasing the likelihood of exploratory behaviour and minimizing the likelihood of aversive reactions (such as neophobic behavioural freezing) in the new environment. Behavioural (reduced exploration; see Methods) and electrophysiological (reduced hippocampal activation; see below) evidence that this type of exploration was accompanied by the acquisition of information about the new environment was found when the animals were reintroduced to the box on the following days.

Experiments were carried out on freely behaving animals that had been habituated over a period of 2 weeks to the recording procedure and the familiar box. Once baseline synaptic transmission, as measured by the amplitude of the field excitatory postsynaptic

potential (EPSP), was found to be stable over a period of at least 3 days, high-frequency conditioning stimulation was applied to the test pathway in order to induce LTP. The conditioning stimulation used in these studies (10 trains of 20 pulses at 200 Hz) was sufficient to elicit a relatively large potentiation of synaptic responses that remained constant over the subsequent 4-h recording period if the animals were kept in the familiar box (see Fig. 1a legend for

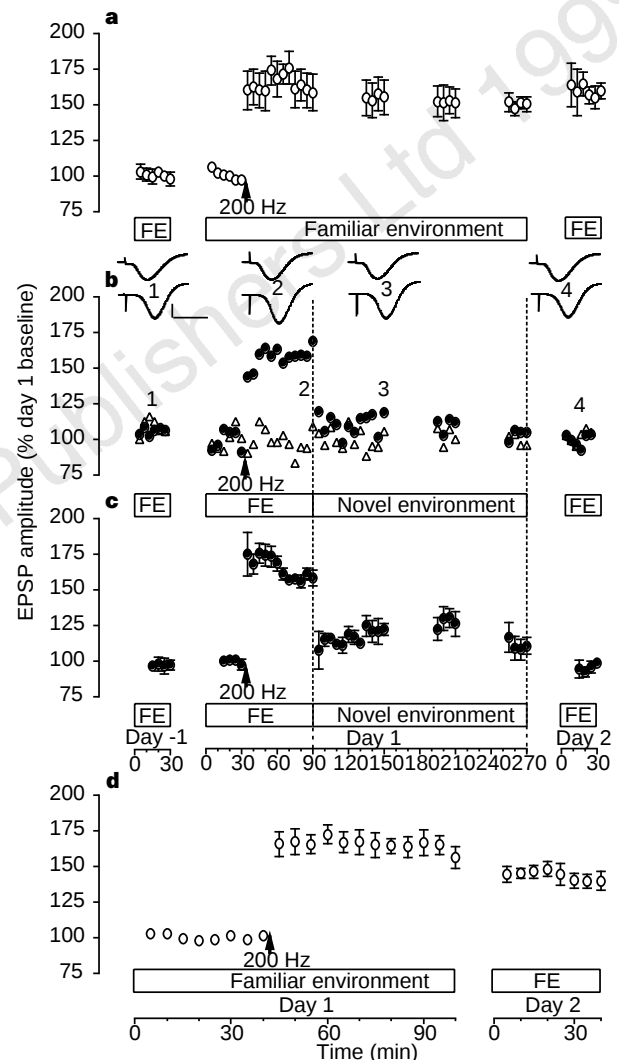


Figure 1 Exploration of a novel environment rapidly reverses LTP. **a**, High-frequency (200 Hz, arrow) stimulation induced stable LTP when induced and recorded in a familiar environment (FE). The amplitude of the field excitatory postsynaptic potential (EPSP) was significantly increased to 158.3 ± 12.9 , 155.4 ± 12 , 151.1 ± 9.9 and $159.7 \pm 5.8\%$ of baseline at 1, 2, 4 and 24 h after the conditioning stimulation (values are 5-min averages $\pm$ s.e.m., $P < 0.01$, $n = 9$). **b**, **c**, LTP was rapidly reversed when the animal was placed in a novel environment 1 h after the application of the high-frequency stimulation. Although the EPSP amplitude was increased at 1 h (159.8 ± 5.4), on introduction to the new environment (NE) synaptic responses returned towards baseline values, reaching $123.8 \pm 4.2\%$ at 2 h and $111.9 \pm 5.9\%$ at 4 h ($P > 0.05$ compared to baseline and $P < 0.01$ compared to potentiated level at 1 h or the level of LTP in controls; $n = 5$). LTP was still absent 24 h later when recorded in the familiar environment ($98.5 \pm 1.7\%$). **b**, Example of a two-pathway experiment. Test (black circles and lower traces) versus ipsilateral non-tetanized control pathway (white triangles and upper traces). Horizontal bar, 10 ms; vertical bar, 2 mV. **d**, Handling the animals by removing them from the familiar box to their home cage 1 h after inducing LTP had no significant effect on the magnitude of LTP when measured 24 h later in the familiar box (156 ± 7.6 and $139.8 \pm 6.3\%$ of baseline at 1 and 24 h after the conditioning stimulation; $P > 0.05$ compared to controls; $n = 5$).

statistics). The LTP was still present on the following day (Fig. 1a) and in some animals persisted for at least a week (data not shown). In marked contrast, when the animals were gently removed from their familiar recording box and placed in the novel box 1 h after the induction of LTP, the synaptic responses rapidly decreased to a level not significantly different from baseline levels (Fig. 1b, c). The reduction in the potentiated responses persisted for the subsequent 3-h recording period in the novel environment, far outlasting the initial increase in general locomotor activity that occurred as the animal explored the new box (<10 min). On the following day, the synaptic responses remained at baseline levels when the recordings were taken 20–40 min after the animals were placed in the familiar box while they were predominantly in a still, alert state (Fig. 1b, c). This confirms that a true reversal of LTP had occurred, rather than a transient state-dependent reduction. There was no change in the synaptic responses in a control pathway in animals that had a second stimulating electrode implanted ipsilaterally (Figs 1b, 2a). Furthermore, in separate control (non-tetanized) pathway experiments, there was no change in basal transmission after rats were moved from the familiar to the novel box ($98.7 \pm 3.9\%$ at 40 min, $P > 0.05$

compared to baseline, $99.4 \pm 3\%$, $n = 5$). Thus, the novel-environment-induced long-lasting reduction in synaptic transmission was restricted to the recently potentiated pathway. The possibility that the mild handling required to move the rats from one box to the other might be responsible for the reversal of LTP was discounted by lifting the animals from the familiar environment and placing them back in their home cage 1 h after the induction of LTP. Even though there was a transient increase in general motor activity on returning to their home cage, similar to that observed in the novel box, there was no significant change in the magnitude of LTP compared to the control animals (Fig. 1d).

An opaque perspex barrier was used to separate the unfamiliar box from the familiar box in subsequent experiments in order to avoid handling the animals at the time of introducing them to the novel environment. When this barrier was removed, the rats freely entered, explored and remained in the novel area. In animals that entered the novel box for the first time 1 h after LTP induction, a rapid reduction of synaptic transmission to baseline levels was observed which persisted while they remained in the novel environment for the next 3 h. This reduction was still present 24 h later

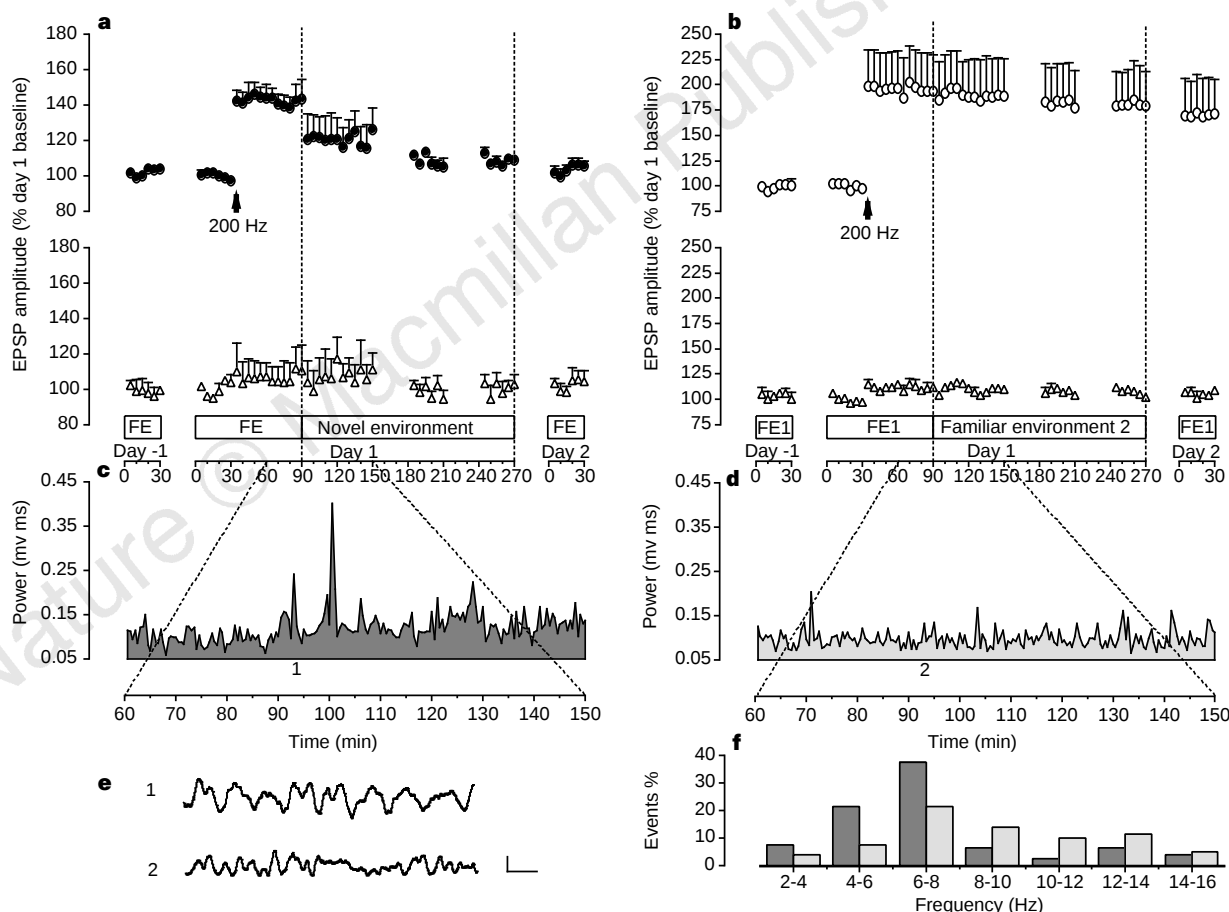


Figure 2 Exploration of a familiar environment fails to affect LTP persistence. **a**, Potentiated responses were reduced in animals that freely left the familiar environment and explored the novel environment for the first time. The test-pathway EPSP decreased from 143.1 ± 11 to $120.4 \pm 12.1\%$ at 1 h after entry into the box (black circles, $P < 0.01$ compared to the potentiated level and $P > 0.05$ compared to baseline; $n = 5$) and stayed at baseline for the rest of the experiment ($105 \pm 2.9\%$ at 24 h). Control pathway responses (white triangles, 107.5 ± 12.4 , 107.9 ± 11.1 and 102.6 ± 3.1 at 1, 2 and 24 h after the tetanus in the test pathway) did not change significantly. **b**, There was no change on entry into the box after a 1-h session of familiarization on the two previous days. LTP was induced 1 h before allowing the rat to cross from the familiar box (familiar environment 1, FE1) to the now familiar 'novel' box (familiar environment 2). There was no significant change in the level of potentiation in the test pathway (circles, 193.7 ± 36.1 ,

188.4 ± 37.7 , 178.6 ± 34.1 and $171.2 \pm 34.2\%$ at 1, 2, 4 and 24 h after the conditioning stimulation; $P > 0.05$, $n = 5$) or in baseline transmission in the control pathway (triangles, 111.5 ± 3.9 , 109.9 ± 3.6 , 102.5 ± 4.4 and $100.6 \pm 6.1\%$ at 1, 2, 4 and 24 h after the conditioning stimulation in the test pathway, $P > 0.05$). **c-f**, First-time, but not third-time, exploration of the box increased the power of the dominant frequency of the hippocampal EEG. **c**, The power increased after the animal explored the box for the first time ($P < 0.01$; 4 animals from **a**). **d**, There was no change after the rat explored the box for the third time ($P > 0.05$; 4 animals from **b**). **e**, Typical traces from **c** and **d**. Horizontal bar, 100 ms; vertical bar, 0.1 mV. **f**, The dominance of 6–8 Hz theta activity increased on first-time exploration (dark bars, $P < 0.05$ compared to third-time exploration, stippled bars, during the first 10-min period after entering the box).

when recordings were taken in the familiar box (Fig. 2a).

We reasoned that if the reversal of LTP was due to experience-dependent hippocampal activation during the exploration of the new box, then re-entry into the same environment after a period of familiarization would not be expected to affect the persistence of recently established LTP. We allowed a group of animals to explore the novel box for a 1-h period on two consecutive days. On the third day, LTP was induced in the familiar box. One hour later, the animals entered and explored the 'novel' box (the second familiar box) for the third time. In these animals, the amplitude of synaptic responses in both the test and control pathways remained stable for the remainder of the experiment (Fig. 2b). The presence of theta electroencephalographic (EEG) activity on entering the novel box was used as an indicator of hippocampal activation^{10–12}. Animals that entered the novel box for the first time had greater theta activity, with a dominant frequency of 6–8 Hz, than those that explored the same box for the third time (Fig. 2c–f). As this difference was observed even when the level of exploratory motor activity was matched between the groups (for example, during the first ~5–10-min period of exploration; Fig. 2c–f), the relative increase in 6–8 Hz theta activity is indicative of an altered hippocampal state associated with the processing of new information, rather than just locomotion^{10–12}. This is strong evidence that it was the process of assimilation of new information about the novel box, rather than some other aspect (such as entry into a darker area), that was responsible for inducing the reversal of LTP.

We also investigated the time window in which exposure to the novel environment was able to affect LTP. In the experiments already described, the change in environment was introduced 1 h after LTP induction, a time when LTP maintenance is considered to be largely independent of protein synthesis¹³. A further set of experiments examined whether the induction of this 'early' phase of LTP was affected by introduction to a novel environment. When the animals were allowed to enter the novel box 5 min before the application of high-frequency stimulation, the stimulation induced an LTP that was stable for at least 1 h after its induction (Fig. 3); thus, the mechanisms necessary for the induction of early-phase LTP were not blocked by the new situation. Another set of experiments studied the effect of introducing the animals to the novel environment 24 h after the induction of LTP, a time when LTP is dependent on both protein and RNA synthesis¹³. In these animals,

exposure to the new environment failed to affect LTP (Fig. 3), indicating that once 'late'-phase LTP has been consolidated, acquisition of information about a new environment had no significant effect on potentiated synaptic responses.

A rapid reversal of LTP ('depotentialization') in the CA1 area of the hippocampus of freely moving adult rats has been reported to be induced by low-frequency stimulation (1–10 Hz) (refs 14–16, but see ref. 17). It is interesting, given the presence of increased theta activity during the induction of naturally occurring reversal of LTP reported here, that the optimal frequencies for artificially inducing depotentialization are in the 5–10-Hz theta frequency range and that stimulation locked to the negative phase of ongoing theta activity is particularly effective^{15,18,19}. During exposure to novelty, hippocampal theta activation can occur synchronously across large populations of hippocampal principal neurons and this synchronization appears to gate processing throughout the hippocampal network in a phase-specific manner¹¹. There is a growing body of evidence that the detection of novelty is important in hippocampal information storage^{20–22}. The time window for the reversal of LTP reported here supports the idea that the hippocampus may hold on to new information until consolidation. Most theories of hippocampal function in memory assume that information is encoded and stored at low density and in a widely distributed manner, thus increasing its storage capacity²³. The finding of complete reversal of recently induced, large-magnitude LTP during exploration of a new environment indicates that extensive experience-dependent reductions in synaptic strength may occur throughout the hippocampal network. Thus the acquisition of new information via the hippocampus may lead to a widespread and complete depotentialization of most recently potentiated synapses, in tandem with a sparsely distributed potentiation of selected synapses. This may explain the failure of this and previous studies to detect long-lasting increases in synaptic strength at so-called 'naive' pathways in the hippocampus following spatial exploration^{4–6}. A redistribution of synaptic efficacy provides a learning mechanism with little net change in synaptic weight²⁴. The active erasure of synaptic efficacy changes, within a defined time after they are initiated, provides protection from lasting effects of inconsequential inputs and endows the system with enhanced combinatorial plasticity²⁵. From our results, reversal of previously established synaptic strengthening may act as a counterpoint to, and presumably together with, the

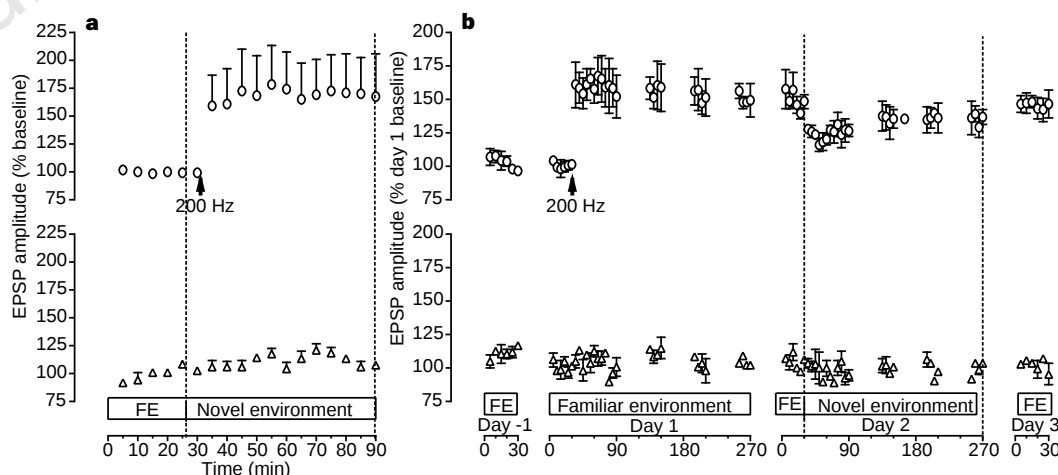


Figure 3 Time window for the effect of novelty exploration on LTP. **a**, LTP induction was not affected by entry into the novel box. When high-frequency (200 Hz, arrow) stimulation was applied to the test pathway 5 min after the animal had entered the novel box, significant LTP was induced (circles, 168 ± 37.8 at 1 h, $P < 0.01$, $n = 5$). Synaptic responses to stimulation of a control input (triangles, $111.7 \pm 3.5\%$, $P > 0.05$) did not change during the experiment. **b**, Well established LTP was not affected by novelty exploration. Stable LTP was induced in the test pathway,

which persisted for 24 h after the high-frequency stimulation (200 Hz, arrow). Allowing the animal to enter freely and explore the novel box at this stage had no effect on synaptic responses either in the test pathway (circles; 152 ± 16 , 148.3 ± 5.3 , 126.7 ± 4.5 , 135.8 ± 11.1 and $146.6 \pm 10.6\%$ of baseline at 1, 24, 25, 28 and 36 h after the conditioning stimulation, $P > 0.05$, $n = 4$) or the control pathway (triangles; 98.7 ± 4.4 , 102.5 ± 2.6 , 97.4 ± 5 , 100.4 ± 2.7 and $103.2 \pm 4.5\%$ at 1, 24, 25, 28 and 36 h after the conditioning stimulation in the test pathway, $P > 0.05$).

strengthening of neuronal connections during the detection and storage of new information by the hippocampus. □

Methods

Electrode implantation and electrophysiology. Experiments were carried out on freely behaving male Wistar rats (200–300 g) that had electrodes implanted under pentobarbitone (60 mg kg⁻¹) anaesthesia. Recordings of field EPSPs were made from the CA1 stratum radiatum of the hippocampus in response to ipsilateral stimulation of the Schaffer collateral/commissural pathway using techniques similar to those described^{12,27}. Animals recovered at least 14 days before the start of the experiment. Test EPSPs were evoked at a frequency of 0.033 Hz and at a stimulation intensity adjusted to give an EPSP amplitude of 50% of maximum. The high-frequency stimulation protocol for inducing LTP consisted of 10 trains of 20 stimuli, interstimulus interval 5 ms (200 Hz), intertrain interval, 2 s. Repeated stimulation with this protocol fails to increase the magnitude of LTP, indicating that it is almost at saturation for the group of synapses under observation²⁶. LTP was measured as mean $\pm$ s.e.m.% of baseline EPSP amplitude recorded over at least a 20-min baseline period. The EEG was simultaneously monitored (from the hippocampal recording electrode) during all experiments so as to ensure that no abnormal activity was evoked by the conditioning stimulation and to monitor hippocampal theta EEG activity. The spectral power of the EEG was measured after fast Fourier transformation of sweeps of 1.2 s duration. Dual pathway experiments, with two independent ipsilateral stimulation inputs to the same recording electrode, were carried out for most experiments. Lack of paired-pulse interaction with responses evoked in the test pathway was used as a criterion of independence.

Recording apparatus and novelty exploration. To allow free exploration without extensive locomotion (which affects brain temperature and field potential measures of synaptic transmission^{4,6,28,29}), the recording boxes were relatively small (0.07 or 0.08 m²). Under these conditions, only very transient (<10 min) and small changes (<1 °C) in brain temperature were observed on entering the novel environment.

Experiments were carried out in a well lit (~750 lux, fluorescent lighting) room. The familiar box was made of clear perspex, whereas the novel box was made of Perspex covered with a thin sheet of plastic which acted as a red filter (>600 nm, filter factor ~3 $\times$). The boxes in the first study had different shapes (34 $\times$ 24 $\times$ 24 cm for the familiar, versus 32 $\times$ 21 $\times$ 20 cm for the novel box). In the other studies, an opaque barrier that separated the familiar and novel environments was removed at 90 min and was closed 20 min later when the animal was in the novel box. To make the novel environment more distinct, the bedding was also different (none in the familiar, versus wood shavings in the novel box). The bedding was changed between rats but was not changed after each trial for a given rat. Behavioural evidence that the animals acquired information about the new environment was provided by the observation that the animals explored less on re-exposure to the novel box on consecutive days (for example, 24 $\pm$ 6 versus 14 $\pm$ 4 transitions between the familiar and novel boxes in the first 20 min on the first and third day, respectively; P < 0.05). Entry into the novel box did not elicit any observable stress responses either hormonally (plasma corticosterone, 5.2 $\pm$ 1.2 versus 3 $\pm$ 0.8 μ g dl⁻¹ in familiar box, measured by HPLC; n = 4)²⁷ or behaviourally (no evidence of behavioural freezing, piloerection or defecation typical of stress). The animals were housed individually in their home cage between recording sessions. Statistical comparisons were made by using Friedman two-way analysis of variance by ranks and Mann–Whitney U -test where appropriate.

Received 20 March; accepted 9 June 1998.

- Martinez, J. & Derrick, B. Long-term potentiation and learning. *Annu. Rev. Psychol.* **47**, 173–203 (1996).
- Jeffery, K. LTP and spatial learning—Where to next? *Hippocampus* **7**, 95–110 (1997).
- Morris, R. G. M. & Frey, U. Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience? *Phil. Trans. R. Soc. Lond. B* **352**, 1489–1503 (1997).
- Hargreaves, E. L., Cain, D. P. & Vandervolf, C. Learning and behavioral long-term potentiation: importance of controlling for motor activity. *J. Neurosci.* **10**, 1472–1478 (1990).
- Erickson, C. A., McNaughton, B. L. & Barnes, C. A. Comparison of long-term enhancement and short-term exploratory modulation of perforant path synaptic transmission. *Brain Res.* **615**, 275–280 (1993).
- Moser, E., Moser, M.-B. & Andersen, P. Potentiation of dentate synapses initiated by exploratory learning in rats: dissociation from brain temperature, motor activity, and arousal. *Learning Memory* **1**, 55–73 (1994).
- Abraham, W. C. & Bear, M. F. Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.* **19**, 126–130 (1996).

- O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
- Eichenbaum, H. Is the rodent hippocampus just for 'place'? *Curr. Opin. Neurobiol.* **6**, 187–195 (1996).
- Green, J. & Arduini, A. Hippocampal electrical activity in arousal. *J. Neurophysiol.* **17**, 553–557 (1954).
- Bland, B. The physiology and pharmacology of hippocampal formation theta rhythms. *Progr. Neurobiol.* **26**, 1–54 (1986).
- Buzsáki, G. Two stage model of memory trace formation: a role for noisy brain states. *Neuroscience* **31**, 551–570 (1989).
- Huang, Y.-Y., Nguyen, P., Abel, T. & Kandel, E. Long-lasting forms of synaptic potentiation in the mammalian hippocampus. *Learning Memory* **3**, 74–85 (1996).
- Stäubli, U. & Lynch, G. Stable depression of potentiated synaptic responses in the hippocampus with 1–5 Hz stimulation. *Brain Res.* **513**, 113–118 (1990).
- Doyle, C. A., Cullen, W. K., Rowan, M. J. & Anwyl, R. Low-frequency stimulation induces homosynaptic depotentiation but not long-term depression of synaptic transmission in the adult anaesthetized and awake rat hippocampus *in vivo*. *Neuroscience* **77**, 75–85 (1997).
- Manahan-Vaughan, D. Group 1 and 2 metabotropic glutamate receptors play differential roles in hippocampal long-term depression and long-term potentiation in freely moving rats. *J. Neurosci.* **17**, 3303–3311 (1997).
- Errington, M. L. *et al.* Stimulation at 1–5 Hz does not produce long-term depression or depotentiation in the hippocampus of the adult rat *in vivo*. *J. Neurophysiol.* **74**, 1793–1799 (1995).
- Huerta, P. & Lisman, J. Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 *in vitro*. *Neuron* **15**, 1053–1063 (1995).
- Stäubli, U. & Chun, D. Factors regulating the reversibility of long-term potentiation. *J. Neurosci.* **16**, 853–860 (1996).
- Parkin, A. J. Human memory: novelty, association and the brain. *Curr. Biol.* **7**, R768–R769 (1997).
- Grunwald, T., Lehnertz, K., Heinze, H. J., Helmstaedter, C. & Elger, C. E. Verbal novelty detection within the human hippocampus proper. *Proc. Natl Acad. Sci. USA* **95**, 3193–3197 (1998).
- Honey, R. C., Watt, A. & Good, M. Hippocampal lesions disrupt an associative mismatch process. *J. Neurosci.* **18**, 2226–2230 (1998).
- Gluck, M. & Myers, C. Psychobiological models of hippocampal function in learning and memory. *Annu. Rev. Psychol.* **48**, 481–514 (1997).
- Miller, K. Synaptic economics: competition and cooperation in synaptic plasticity. *Neuron* **17**, 371–374 (1996).
- Dudai, Y. Consolidation: fragility on the road to the engram. *Neuron* **17**, 367–370 (1996).
- Doyle, C., Holscher, C., Rowan, M. J. & Anwyl, R. The selective neuronal NO synthase inhibitor 7-nitro-indazole blocks both long-term potentiation and depotentiation of field EPSPs in rat hippocampal CA1 *in vivo*. *J. Neurosci.* **16**, 418–424 (1996).
- Xu, L., Anwyl, R. & Rowan, M. J. Behavioural stress facilitates the induction of long-term depression in the hippocampus. *Nature* **387**, 497–500 (1997).
- Leung, L. Behavior-dependent evoked potentials in the hippocampal CA1 region of the rat. I. Correlation with behavior and EEG. *Brain Res.* **198**, 95–117 (1980).
- Moser, E., Mathiesen, I. & Anderson, P. Association between brain temperature and dentate field potentials in exploring and swimming rats. *Science* **259**, 1324–1326 (1993).

Acknowledgements. This research was supported by the Health Research Board of Ireland, the European Union DGXII and the Wellcome Trust. We thank W. K. Cullen and J. Wu for assistance.

Correspondence and requests for material should be addressed to M.J.R. (e-mail: mrowan@mail.tcd.ie).

Decreased lesion formation in CCR2^{-/-} mice reveals a role for chemokines in the initiation of atherosclerosis

Landin Boring^{*†}, Jennifa Gosling^{*}, Michael Cleary^{*} & Israel F. Charo^{*‡}

^{*} Gladstone Institute of Cardiovascular Disease, San Francisco, California 94141, USA

[†] Cardiovascular Research Institute, University of California, San Francisco, California 94110, USA

[‡] Department of Medicine and the Daiichi Research Center, University of California, San Francisco, California 94110, USA

Chemokines are proinflammatory cytokines that function in leukocyte chemoattraction and activation and have recently been shown to block the HIV-1 infection of target cells through interactions with chemokine receptors^{1,2}. In addition to their function in viral disease, chemokines have been implicated in the pathogenesis of atherosclerosis. Expression of the CC chemokine monocyte chemoattractant protein-1 (MCP-1) is upregulated in human atherosclerotic plaques^{3,4}, in arteries of primates on a hypercholesterolaemic diet⁵ and in vascular endothelial and smooth muscle cells exposed to minimally modified lipids^{5,6}. To determine whether MCP-1 is causally related to the development of atherosclerosis, we generated mice that lack CCR2, the receptor for MCP-1 (ref. 7), and crossed them with apolipoprotein (apo) E-

null mice⁸⁻¹⁰ which develop severe atherosclerosis. Here we show that the selective absence of CCR2 decreases lesion formation markedly in apoE^{-/-} mice but has no effect on plasma lipid or lipoprotein concentrations. These data reveal a role for MCP-1 in the development of early atherosclerotic lesions and suggest that upregulation of this chemokine by minimally oxidized lipids is an important link between hyperlipidaemia and fatty streak formation.

The arterial fatty streak is composed of lipid-laden macrophages (foam cells) and is the precursor of more complex and dangerous lesions. The molecular signals that initiate monocyte/macrophage recruitment to the vessel wall are, however, unknown. We stained aortic root sections from mice on the Western diet¹¹ with MOMA-2, a macrophage-specific antibody¹². Macrophages were abundant in the subendothelial space of CCR2^{+/+}apoE^{-/-} animals fed the high-fat diet for as little as 5 weeks, and constituted most of the cells in the lesion (Fig. 1a). In contrast, markedly fewer macrophages were present in the aortas of CCR2^{-/-}apoE^{-/-} mice (Fig. 1b). Quantitative analysis revealed significantly less MOMA-2-positive staining in CCR2^{-/-} mice than in CCR2^{+/+} mice (Fig. 1c). These data indicate that activation of CCR2 was important in recruitment of mono-

cytes/macrophages into the vessel wall, the earliest recognizable sign of atherosclerosis. Staining with Oil Red O confirmed the significant decrease in lesion area in CCR2^{-/-} mice fed the Western diet for 5 weeks (Fig. 2). Lesion size in the CCR2^{+/-} mice was intermediate between the lesion sizes of wild-type and CCR2^{-/-} mice, suggesting a gene dosage effect (Fig. 2). After 9 weeks on the diet, there was no difference between the CCR2^{+/-} and wild-type mice, and both had significantly larger lesions than the CCR2^{-/-} mice. The reduction in lesion size was observed throughout the aortic root and was most pronounced in the valve leaflet region, where wild-type mice developed the most severe lesions (not shown). After 13 weeks on the diet, CCR2^{-/-} mice still had significantly smaller lesions than CCR2^{+/-} mice (Fig. 2). Thus, lesion development was decreased at all time points examined. In addition, at both 9 and 13 weeks, the lesions were less complex in the CCR2^{-/-} mice (not shown). These data suggest that the decreased recruitment of macrophages observed at 5 weeks reduced lesion size at the later time points.

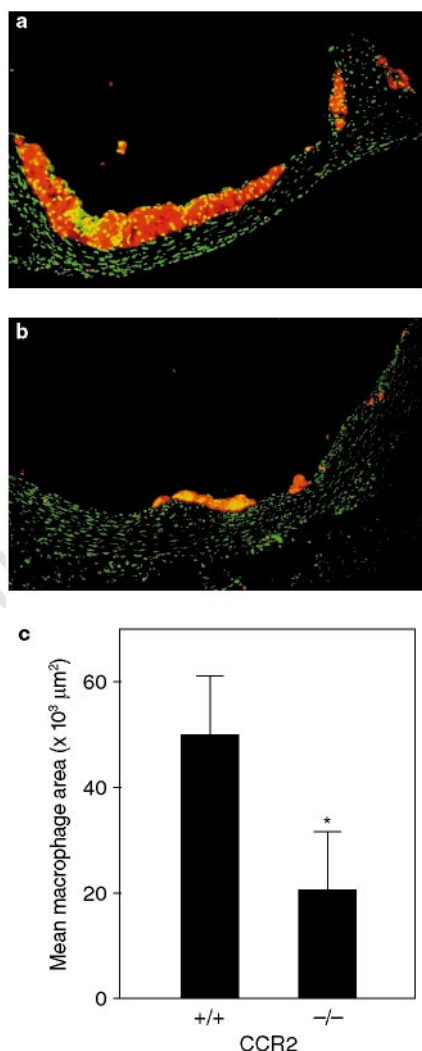


Figure 1 Macrophage infiltration of the aortic sinus in apoE^{-/-} mice fed a high-fat (Western) diet for 5 weeks. **a**, CCR2^{+/+}, **b**, CCR2^{-/-}. Sections from the aortic sinus were stained for macrophages with MOMA-2 (red) and for nuclei with SYTOX green. Regions of overlap appear yellow. Shown are representative sections from mice of each genotype. Original magnification $\times 125$. **c**, Quantitation of MOMA-2 staining. Values are means $\pm$ s.d. ($n =$ six mice of each genotype). $P = 0.0043$ for CCR2^{+/+} versus CCR2^{-/-} (Mann-Whitney test).

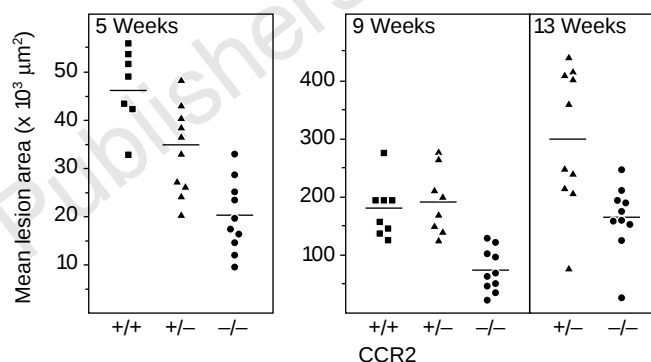


Figure 2 Mean lesion area in the aortic root. Cross-sections ($10 \mu\text{m}$) of the aorta were stained for lipid with Oil Red O and quantitated by digital morphometry. Each symbol represents one animal; bars represent means. Spearman rank order analysis suggested a gene dosage effect at the 5-week time point ($r = 0.8029$, $P < 0.0001$). For comparisons between groups, $P < 0.001$ for CCR2^{+/+} versus CCR2^{-/-} and $P < 0.05$ for CCR2^{+/-} versus CCR2^{-/-} at 5 weeks (ANOVA); $P < 0.01$ for CCR2^{+/+} versus CCR2^{-/-} and $P < 0.001$ for CCR2^{+/-} versus CCR2^{-/-} at 9 weeks (ANOVA); $P = 0.0029$ for CCR2^{+/-} versus CCR2^{-/-} at 13 weeks (Mann-Whitney test).

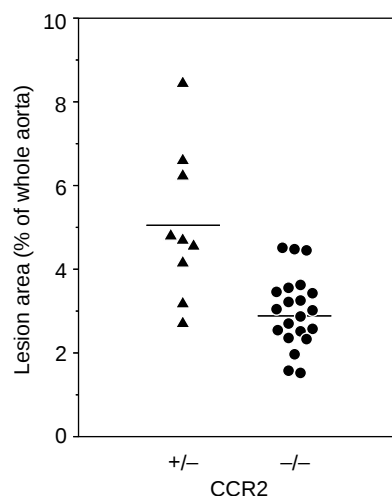


Figure 3 Lesion coverage in the aortas of CCR2^{+/+} and CCR2^{-/-} mice. Whole aortas from mice fed the high-fat diet for 13 weeks were mounted *en face* and stained for lipid with Sudan IV. Total lesion area from the aortic sinus to the femoral bifurcation was quantitated by digital morphometry. Each symbol represents one animal; bars represent means. The percentage of the aortic surface covered by lesions in the CCR2^{-/-} mice ($n = 21$) was significantly lower than in the CCR2^{+/+} mice ($n = 9$) ($P = 0.001$, Mann-Whitney test).

Table 1 Total cholesterol in serum from mice on the Western diet

Genotype	Total cholesterol (mg dl ⁻¹)		
	5 weeks	9 weeks	13 weeks
CCR2 ^{+/+} , apoE ^{-/-}	1916 ± 291 (17)	1947 ± 376 (8)	ND
CCR2 ^{-/-} , apoE ^{-/-}	1809 ± 408 (20)	1963 ± 354 (11)	1976 ± 402 (10)
CCR2 ^{-/-} , apoE ^{-/-}	1661 ± 431 (17)	1851 ± 513 (11)	1788 ± 375 (21)

Values shown are means ± s.d. (numbers of mice). ND, no data. Cholesterol levels were not significantly different between genotypes. $P = 0.1062$ for 5 weeks and $P = 0.8060$ for 9 weeks by ANOVA; $P = 0.1832$ for 13 weeks by Mann-Whitney test.

We next examined lesion development along the entire aorta in whole-mount *en face* preparations stained with Sudan IV. In animals on the Western diet for 5 and 9 weeks, less than 1% of the aorta was covered with lesions, and no significant differences between the CCR2 genotypes were observed (data not shown). At 13 weeks, however, lesions covered 5.1% of the aorta in the CCR2^{+/+} mice, but only 2.9% in the CCR2^{-/-} mice (Fig. 3). Thus, by two independent techniques, lesion size was significantly reduced in the absence of CCR2. Longer duration studies will be required to determine whether the size and complexity of the lesions in the CCR2^{-/-} mice eventually equal those of the wild-type animals.

To determine whether differences in lipid metabolism could account for the decreased atherosclerosis in the CCR2^{-/-} mice, we examined plasma lipid levels and lipoprotein profiles. Total serum cholesterol levels were not statistically different between the CCR2 genotypes at any time point and were within the range expected for apoE^{-/-} mice maintained on this diet¹¹ (Table 1). In addition, no differences were observed in plasma triglyceride levels (285.6 ± 107.7 mg dl⁻¹, $n = 6$ for CCR2^{+/+}; 282.8 ± 86.4 mg dl⁻¹, $n = 5$ for CCR2^{-/-}; 307.2 ± 79.6 mg dl⁻¹, $n = 7$ for CCR2^{-/-}). As analysed by gel filtration chromatography, the plasma lipoprotein profiles were essentially identical in wild-type and CCR2^{-/-} mice (Fig. 4). Consistent with earlier reports, most of the cholesterol in apoE^{-/-} mice was contained within very-low-density lipoprotein-sized particles⁸. Lesion size in apoE^{-/-} mice can be influenced by high-density lipoprotein (HDL) levels¹⁰, but HDL cholesterol levels were very low and were unaffected by the CCR2 genotype (Fig. 4). These data suggest that the decreased lesion size in the CCR2^{-/-} mice was not accounted for by changes in plasma cholesterol levels or in the distribution of lipoprotein particles. Gupta *et al.*¹³ reported decreased atherosclerosis in interferon- γ receptor-deficient mice on the apoE^{-/-} background and also noted an increase in plasma levels of apoA-IV, a potentially atheroprotective lipoprotein¹⁴. We found no difference in plasma apoA-IV concentrations between CCR2^{+/+} and CCR2^{-/-} mice (data not shown). The CCR2^{-/-} mice were backcrossed with C57Bl/6 mice, resulting in study animals that

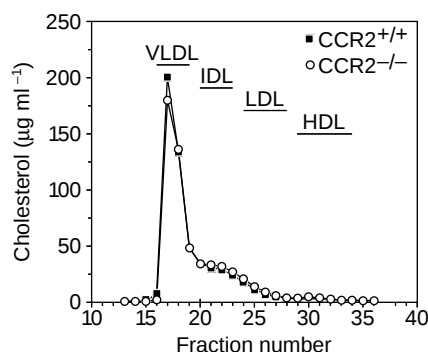


Figure 4 Lipoprotein profile of plasma from CCR2^{+/+} and CCR2^{-/-} mice. Values are averages of four mice of each genotype with similar total cholesterol levels. There were no significant differences in total non-HDL cholesterol (CCR2^{+/+}, $1,664 \pm 151$ mg dl⁻¹; CCR2^{-/-}, $1,510 \pm 210$ mg dl⁻¹) or HDL cholesterol (CCR2^{+/+}, 49 ± 27 mg dl⁻¹; CCR2^{-/-}, 54 ± 3.4 mg dl⁻¹) (Mann-Whitney). LDL, low-density lipoprotein; IDL, intermediate-density lipoprotein.

were 87.5% C57Bl/6 and 12.5% 129/Sv. Although unlikely, the possibility that a 129/Sv gene contributed to the differences in atherosclerosis cannot be completely excluded, particularly if this gene was linked to CCR2.

In summary, these studies identify CCR2 as a genetic determinant of murine atherosclerosis and provide strong evidence for a direct non-cholesterol-mediated effect of MCP-1 in macrophage recruitment and atherogenesis. These results are consistent with those of Suzuki *et al.*¹⁵, who reported a role for the macrophage scavenger receptor in atherosclerosis susceptibility. Monocytes also express CCR1 and CCR5, and atherosclerosis studies of mice in which multiple receptors have been genetically deleted will probably be of great interest.

Methods

Genetically modified mice. CCR2^{-/-} mice⁷ (hybrids of C57Bl/6 and 129/SvJae) were mated with apoE^{-/-} mice⁹ on the C57Bl/6 background (N10 C57Bl/6, Jackson Labs). CCR2^{+/+}, apoE^{-/-} mice were backcrossed with apoE^{-/-} mice to produce CCR2^{+/+}, apoE^{-/-} mice; these mice were intercrossed to produce all three CCR2 genotype on the apoE-null background (87.5% C57Bl/6; 12.5% 129/SvJae). The double-knockout (apoE^{-/-}, CCR2^{-/-}) mice were born at the expected Mendelian ratios, developed normally, and were disease-free while maintained in a pathogen-free environment. Mice were weaned at 4 weeks, fed normal rodent chow (4.5% fat; Ralston Purina Co.) for an additional week, and switched to the Western diet (21% fat, 0.15% cholesterol; Harlan Tekland no. 88137) at 5 weeks of age¹¹.

Quantitation of atherosclerotic lesions. Hearts and aortas were prepared essentially as described^{16,17}, except that mice were perfused with 3.0% paraformaldehyde in PBS and tissues were equilibrated in 20% sucrose before embedding in OCT compound (Tissue Tek). Atherosclerosis was assessed by two independent techniques: Oil Red O staining of lesions in cross-sections from the aortic root and Sudan IV staining of lesions in pinned-out aortas. Cryosections (10 μ m) spanning 550 μ m of the proximal aorta were collected, and every fifth section, extending 250 μ m in both directions from the coronary artery branch point, was stained with Oil Red O. Lesion areas were quantified with a digital colour video camera and Image-1/AT software as described¹⁶. Adjacent sections were stained with haematoxylin and eosin for morphological analysis. The remainder of the aorta was opened longitudinally along the ventral midline from the aortic root to the iliac arteries, and the lesion area in *en face* preparations stained with Sudan IV was quantitated as described¹⁷. Image analysis was performed by a trained observer blinded to the genotype of the mice.

Immunohistochemistry. Serial cryosections collected at 50- μ m intervals were stained with a monoclonal rat antibody to the mouse monocyte-macrophage marker MOMA-2 (Biosource International, Camarillo, CA; 1:400 dilution), followed by detection with biotinylated secondary antibodies and streptavidin-horseradish peroxidase (PharMingen, San Diego). The signal was enhanced with the NEN DuPont Tyramide Signal Amplification kit, and sections were counterstained for nuclei with SYTOX green (Molecular Probes, Eugene, OR). Fluorescent images were captured into Image-1 as described above. Data represent the average area of MOMA-2 staining from 6–8 sections per mouse.

Lipid analysis. Cholesterol and triglyceride concentrations were determined with colorimetric assay kit (Spectrum cholesterol kit, Abbott Labs; GPO triglyceride kit, Boehringer Mannheim). Cholesterol levels were measured in serum of nonfasted mice collected at the time of killing, and triglycerides were assayed in plasma obtained after a 4-h fast. For lipoprotein analysis, plasma was prepared from blood collected by retro-orbital puncture into tubes containing EDTA (1 mM) and aprotinin (113×10^3 U ml⁻¹, ICN). Plasma samples from individual mice (50 μ l in a total volume of 300 μ l) were fractionated by FPLC on a Superpose 6 column (HR 10/30, Pharmacia LKB) as described¹⁶. ApoA-IV levels in plasma were determined by gradient gel electrophoresis (SDS-PAGE, 4–15% Tris-HCl Ready Gel; Bio-Rad) of very-low-density lipoprotein (fractions 17–19; Fig. 4) from four CCR2^{+/+} and four CCR2^{-/-} mice, followed by western blot analysis with a rabbit polyclonal antibody against rat apoA-IV (generously provided by Karl Weisgraber, Gladstone Institute of Cardiovascular Disease).

Statistical analysis. For analysis of lesion size, comparisons between groups were performed using the Kruskal-Wallis nonparametric ANOVA or Mann-

Whitney *U* test. The Spearman rank order correlation was used to analyse the relationship between CCR2 gene dosage and lesion area. Instat 2.01 software for the Macintosh was used for all calculations.

Received 11 May; accepted 3 July 1998.

- Cocchi, F. *et al.* Identification of RANTES, MIP-1 α , and MIP-1 β as the major HIV-suppressive factors produced by CD8⁺ T cells. *Science* **270**, 1811–1815 (1995).
- Choe, H. *et al.* The β -chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**, 1135–1148 (1996).
- Nelken, N. A., Coughlin, S. R., Gordon, D. & Wilcox, J. N. Monocyte chemoattractant protein-1 in human atherosclerotic plaques. *J. Clin. Invest.* **88**, 1121–1127 (1991).
- Ylä-Herttuala, S. *et al.* Expression of monocyte chemoattractant protein 1 in macrophage-rich areas of human and rabbit atherosclerotic lesions. *Proc. Natl Acad. Sci. USA* **88**, 5252–5256 (1991).
- Yu, X. *et al.* Elevated expression of monocyte chemoattractant protein 1 by vascular smooth muscle cells in hypercholesterolemic primates. *Proc. Natl Acad. Sci. USA* **89**, 6953–6957 (1992).
- Cushing, S. D. *et al.* Minimally modified low density lipoprotein induces monocyte chemotactic protein 1 in human endothelial cells and smooth muscle cells. *Proc. Natl Acad. Sci. USA* **87**, 5134–5138 (1990).
- Boring, L. *et al.* Impaired monocyte migration and reduced type 1 (Th1) cytokine responses in C-C chemokine receptor 2 knockout mice. *J. Clin. Invest.* **100**, 2552–2561 (1997).
- Plump, A. S. *et al.* Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell* **71**, 343–353 (1992).
- Zhang, S. H., Reddick, R. L., Piedrahita, J. A. & Maeda, N. Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science* **258**, 468–471 (1992).
- Plump, A. S., Scott, C. J. & Breslow, J. L. Human apolipoprotein A-I gene expression increases high density lipoprotein and suppresses atherosclerosis in the apolipoprotein E-deficient mouse. *Proc. Natl Acad. Sci. USA* **91**, 9607–9611 (1994).
- Nakashima, Y., Plump, A. S., Raines, E. W., Breslow, J. L. & Ross, R. ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree. *Arterioscler. Thromb.* **14**, 133–140 (1994).
- Kraal, G., Rep, M. & Janse, M. Macrophages in T and B cell compartments and other tissue macrophages recognized by monoclonal antibody MOMA-2. An immunohistochemical study. *Scand. J. Immunol.* **26**, 653–661 (1987).
- Gupta, S. *et al.* IFN- γ potentiates atherosclerosis in apoE knock-out mice. *J. Clin. Invest.* **99**, 2752–2761 (1997).
- DuVerger, N. *et al.* Protection against atherogenesis in mice mediated by human apolipoprotein A-IV. *Science* **273**, 966–968 (1996).
- Suzuki, H. *et al.* A role for macrophage scavenger receptors in atherosclerosis and susceptibility to infection. *Nature* **386**, 292–296 (1997).
- Purcell-Huynh, D. A. *et al.* Transgenic mice expressing high levels of human apolipoprotein B develop severe atherosclerotic lesions in response to a high-fat diet. *J. Clin. Invest.* **95**, 2246–2257 (1995).
- Véniant, M. M. *et al.* Susceptibility to atherosclerosis in mice expressing exclusively apolipoprotein B48 or apolipoprotein B100. *J. Clin. Invest.* **100**, 180–188 (1997).

Acknowledgements. We thank D. Sanan, D. Newland and L. Jensen for help with tissue preparation and staining; D. Dichek for advice on statistical analysis; J. C. W. Carroll, N. C. Shea and S. Gonzales for figure preparation; J. Ernst for assistance with fluorescent microscopy; G. Howard and S. Ordway for editorial expertise; and A. Chen for manuscript preparation. We also thank R. Pitas, R. Farese and K. Weisgraber for careful readings of the manuscript. This work was supported in part by a grant from the NIH to I.E.C.

Correspondence and requests for materials should be sent to I.E.C. at the Gladstone Institute of Cardiovascular Disease (e-mail: icharo@gladstone.ucsf.edu).

Leptin modulates the T-cell immune response and reverses starvation-induced immunosuppression

Graham M. Lord^{*†}, Giuseppe Matarese^{*†‡}, Jane K. Howard^{†§}, Richard J. Baker^{*}, Stephen R. Bloom[§] & Robert I. Lechler^{*}

Imperial College School of Medicine, Departments of ^{*}Immunology and [§]Endocrinology, The Hammersmith Hospital, Du Cane Road, London W12 0NN, UK

[†] These authors contributed equally to this work.

Nutritional deprivation suppresses immune function^{1–3}. The cloning of the *obese* gene and identification of its protein product leptin⁴ has provided fundamental insight into the hypothalamic regulation of body weight^{5,6}. Circulating levels of this adipocyte-derived hormone are proportional to fat mass^{6,7} but may be lowered rapidly by fasting^{8,9} or increased by inflammatory mediators^{10,11}. The impaired T-cell immunity of mice^{12,13} now known to be

defective in leptin (*ob/ob*)⁴ or its receptor (*db/db*)^{14,15}, has never been explained. Impaired cell-mediated immunity^{1–3} and reduced levels of leptin⁷ are both features of low body weight in humans. Indeed, malnutrition predisposes to death from infectious diseases¹⁶. We report here that leptin has a specific effect on T-lymphocyte responses, differentially regulating the proliferation of naive and memory T cells. Leptin increased Th1 and suppressed Th2 cytokine production. Administration of leptin to mice reversed the immunosuppressive effects of acute starvation. Our findings suggest a new role for leptin in linking nutritional status to cognate cellular immune function, and provide a molecular mechanism to account for the immune dysfunction observed in starvation.

Most immune responses are orchestrated by CD4⁺ helper T cells (Th). We first determined the effect of leptin on Th responses in the context of the mixed-lymphocyte reaction (MLR) resulting from the culture of T cells with major histocompatibility complex (MHC)-incompatible (allogeneic) stimulator cells. The doses of leptin used in these experiments were chosen to incorporate the range of serum levels measured in humans¹⁷. Leptin induced a

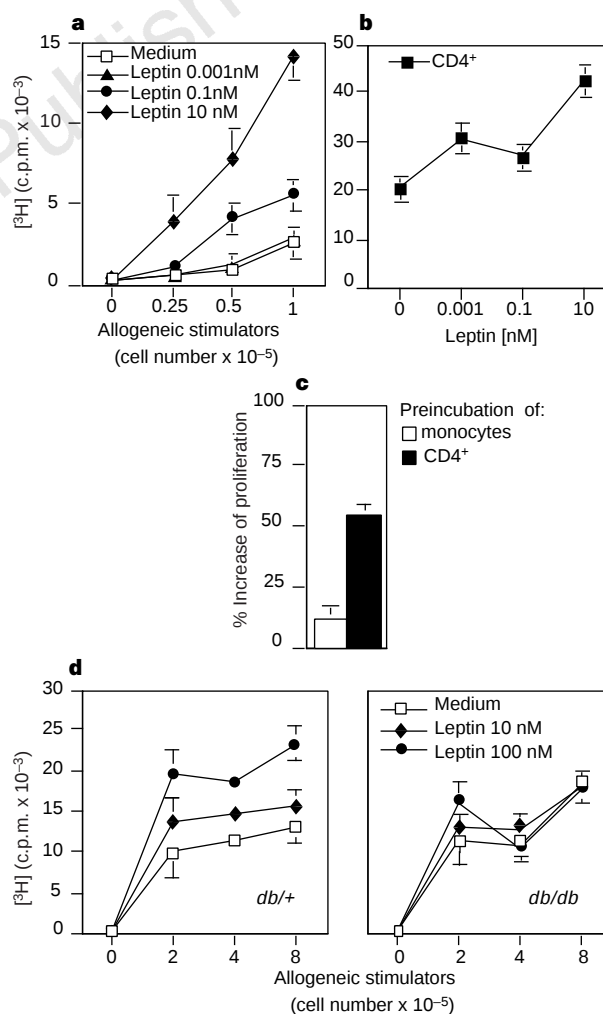


Figure 1 Leptin enhances the alloproliferative response. **a**, **b**, Thymidine incorporation in the presence of leptin in an MLR performed using human PBLs as responder and MHC-mismatched PBMCs as stimulator cells (**a**) or highly purified CD4⁺ T cells as responders as a responder/stimulator ratio of 1:1 (**b**) (6 experiments). **c**, Preincubation of either responder CD4⁺ T cells or irradiated stimulator allogeneic monocytes with 10 nM leptin before co-culture in the absence of leptin. **d**, MLR using murine heterozygous *db*^{+/+} (H-2^d) and homozygous *db*^{/db} (H-2^b) splenocytes as responders with irradiated C57BL/6 (H-2^b) allogeneic splenocytes (3 experiments). All data are expressed as mean c.p.m. of triplicate cultures $\pm$ s.e.m.

[‡] Present address: Laboratorio di Immunologia, Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università di Napoli "Federico II", via Pansini 5, 80131, Napoli, Italy.

marked dose-dependent increase in the proliferative response of peripheral blood lymphocytes (PBLs) in response to allogeneic peripheral blood mononuclear cells (PBMCs) (Fig. 1a). The alloproliferative response was augmented in a similar way when the experiment was repeated using highly purified CD4⁺ cells as responders, suggesting that T cells, rather than accessory cells in the responder population, were the target of leptin action (Fig. 1b). To determine whether it was the responder or the stimulator cells that were being affected by leptin, either responder or stimulator populations were preincubated for 12 hours in the presence or absence of leptin and then washed extensively before co-culture in the absence of leptin. Preincubation of the responder CD4⁺ T cells with leptin produced a greater enhancement of proliferation in the subsequent MLR than did preincubation of the stimulator cells (Fig. 1c). The augmentation of proliferation was still seen, despite the fact that the leptin was only present for 12 hours during preincubation, rather than for the duration of the MLR (5 days) as in the previous experiments. Incubation of CD4⁺ T cells with leptin, in the absence of allogeneic stimulator cells, produced no increase in proliferation (data not shown), suggesting that cognate recognition by the T-cell antigen receptor (TCR) was required before leptin could have its effect.

To confirm the specificity of these effects, we compared the responses of T cells from leptin receptor-defective *db/db* mice with those of their littermate controls (*db/+*). Several splice variants of the leptin receptor (ObR) are expressed *in vivo*. Although all of the receptor isoforms bind leptin, the long isoform of the receptor (ObRb) appears to be of prime importance in signal transduction^{14,15}. A point mutation within the ObR gene of the *diabetic (db)* mouse generates a new splice donor site that dramatically reduces the expression of the long isoform in homozygous *db/db* mice^{14,15} and renders them resistant to the weight-lowering effects of endogenous and exogenous leptin⁵. The alloproliferative response of splenocytes from homozygous *db/db* mice was similar to that of heterozygous *db/+* mice in the absence of leptin, consistent with reports that these cells can function normally *ex vivo*¹². However, the leptin-mediated augmentation of proliferation in response to allogeneic splenocytes seen with responder cells from *db/+* mice was not seen with *db/db* responder cells, despite the higher doses of leptin used (Fig. 1d). Similar results were obtained when MHC class II-transfected murine fibroblasts were used as allogeneic stimulators. All the MLR responses were inhibited by addition of an appropriate anti-MHC class II antibody (data not shown). These results indicate that the observed enhancement of proliferation by leptin was due to a specific effect of leptin receptor signalling, rather than to a nonspecific mitogenic stimulus.

Using long-isoform-specific primers, we were able to detect the expression of ObRb messenger RNA by the reverse transcriptase-polymerase chain reaction (RT-PCR) in human hypothalamus and PBMC, but not in kidney, in agreement with previous reports^{15,18}. Consistent with our functional data, we detected ObRb mRNA in

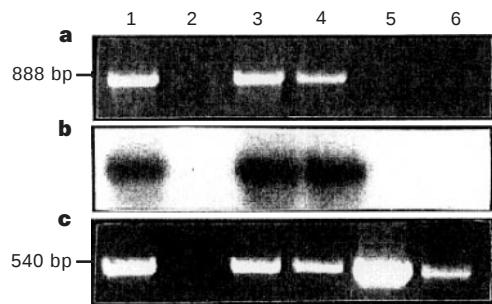


Figure 2 The long isoform of the leptin receptor is expressed in CD4⁺ T cells. **a**, RT-PCR with human ObRb specific primers: lane 1, hypothalamus; lane 2, no RNA control; lane 3, PBMCs; lane 4, CD4⁺ T cells; lane 5, monocytes; lane 6, renal mesangial cells. **b**, Southern blot analysis to confirm the specificity of the PCR product. **c**, RT-PCR with primers specific for β-actin.

highly purified resting human CD4⁺ T cells but not in freshly isolated monocytes (Fig. 2). Taken together, these results indicate that leptin enhances T-cell responses primarily by binding to its receptor on the T cells, rather than through a direct effect on the stimulator cell.

The MLR provokes a strong proliferative response by both naive and memory T cells¹⁹. In time-course studies, the effect of leptin in an MLR was maximal at day 7, a kinetic characteristic of a primary (naive) immune response (data not shown). We therefore investigated whether leptin had differential effects on naive (CD45RA⁺) and memory (CD45RO⁺) CD4⁺ T cells, which both express ObRb mRNA (not shown). Depletion of CD45RO⁺ rather than CD45RA⁺ cells from the responding population of PBMCs in an MLR enhanced the response to leptin (Fig. 3a). Proliferative responses were also determined in an MLR after preincubating highly purified naive or memory T cells for 12 hours either in the presence or absence of leptin, and then washing the cells extensively before culturing them with allogeneic stimulator cells. The effect of leptin was much more pronounced on the CD4⁺ CD45RA⁺ T cells than on CD4⁺ CD45RO⁺ T cells (Fig. 3b). Further evidence that leptin amplifies primary T-cell responses was provided by the use of responder T cells from umbilical cord blood (UCB), the purest preparation of naive

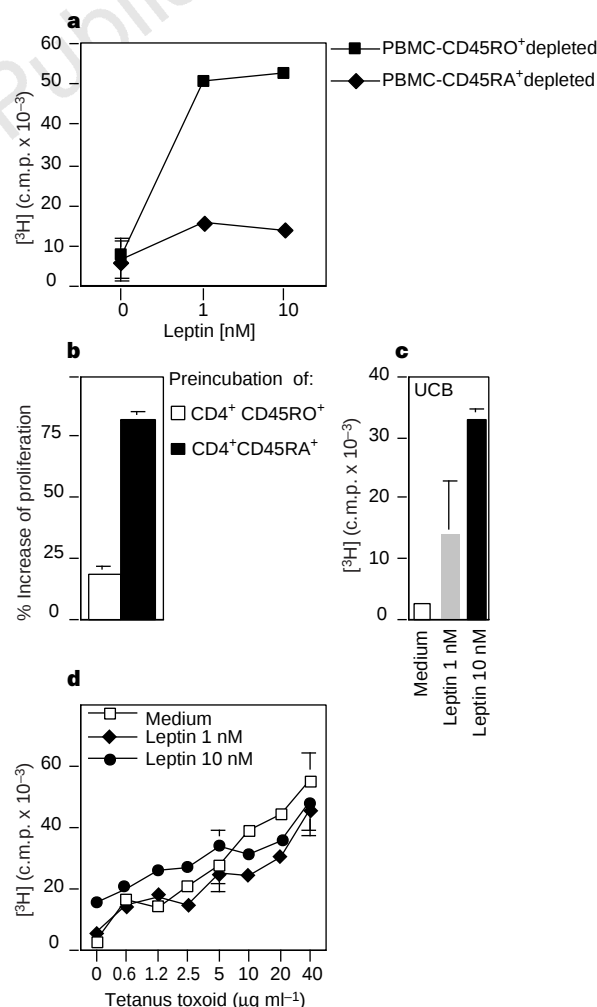


Figure 3 Leptin differentially affects the proliferative responses of naive and memory T cells. **a**, MLR with either human CD45RO⁺ or CD45RA⁺ depleted PBMCs as responders. **b**, Preincubation with leptin (10 nM) of either CD4⁺ CD45RA⁺ T cells or CD4⁺ CD45RO⁺ T cells before an MLR in the absence of leptin (3 experiments). **c**, MLR using human umbilical cord blood (UCB) responder cells (6 experiments). **d**, An adult T-cell recall antigen response to tetanus toxoid (3 experiments). All data are expressed as mean c.p.m. of triplicate cultures $\pm$ s.e.m.

T cells. As shown in Fig. 3c, proliferation was increased tenfold in the presence of leptin. The reduced effect of leptin on proliferation by memory phenotype T cells in the MLR was paralleled by the finding that leptin had little effect on recall responses of previously immunized individuals to tetanus toxoid (Fig. 3d), or on proliferation of a panel of established human T-cell clones (data not shown).

CD4⁺ T cells (Th) determine the nature of immune effector function according to the pattern of cytokines that they secrete²⁰.

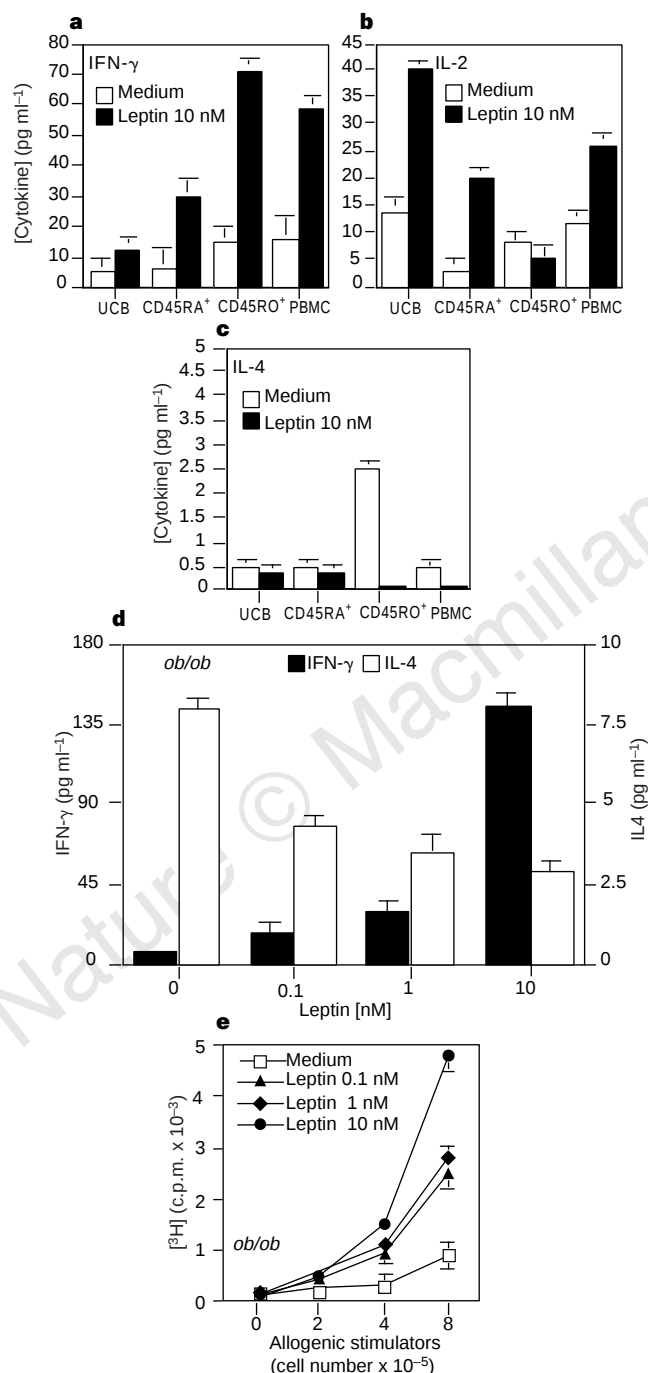


Figure 4 Leptin favours the production of proinflammatory cytokines. **a–c**, Cytokine production in the presence or absence of leptin in an MLR using human UCB cells, naive (CD45RA⁺), memory (CD45RO⁺) T cells or PBMCs as responders: **a**, Interferon- γ ; **b**, interleukin-2; **c**, interleukin-4. **d**, Interferon- γ and interleukin-4 production by *ob/ob* T cells in an MLR in the presence of increasing doses of leptin. **e**, Proliferation of *ob/ob* responder T cells in an MLR in the presence of increasing doses of leptin, done in autologous *ob/ob* serum. All data are expressed as mean $\pm$ s.e.m.

This cell population can be divided into two subsets: Th1 cells, which secrete proinflammatory cytokines (such as interleukin(IL)-2 and interferon(IFN)- γ), and Th2 cells, which secrete cytokines with predominantly regulatory functions (such as IL-4). Several factors have been identified that polarize Th cells into these subsets²⁰. We tested the possibility that leptin influences Th cytokine production by measuring IL-2, IFN- γ and IL-4 in MLR experiments (Fig. 4). We found that leptin increased IL-2 production by all the T-cell populations studied apart from the CD45RO⁺ cells, which, despite a lack of proliferation and IL-2 production, showed a substantial increase in IFN- γ secretion (Fig. 4a, b). A similar increase in IFN- γ secretion was seen with the other T-cell preparations. Only the CD45RO⁺ population secreted measurable amounts of IL-4, and this was completely inhibited by the addition of leptin (Fig. 4c). These results indicate that leptin may bias T-cell responses towards a proinflammatory phenotype (Th1). We therefore analysed the cytokine and alloproliferative responses of responder cells from C57BL/6 *ob/ob* mice. This mouse strain is unable to produce functional leptin, and therefore its T cells have never been exposed to leptin *in vivo*⁴. In the absence of leptin, *ob/ob* T cells produced minimal IFN- γ and a moderate amount of IL-4 (Fig. 4d). The addition of leptin induced the production of large amounts of IFN- γ and suppression of IL-4 production in a dose-dependent manner. Leptin increased both proliferation (Fig. 4e) and

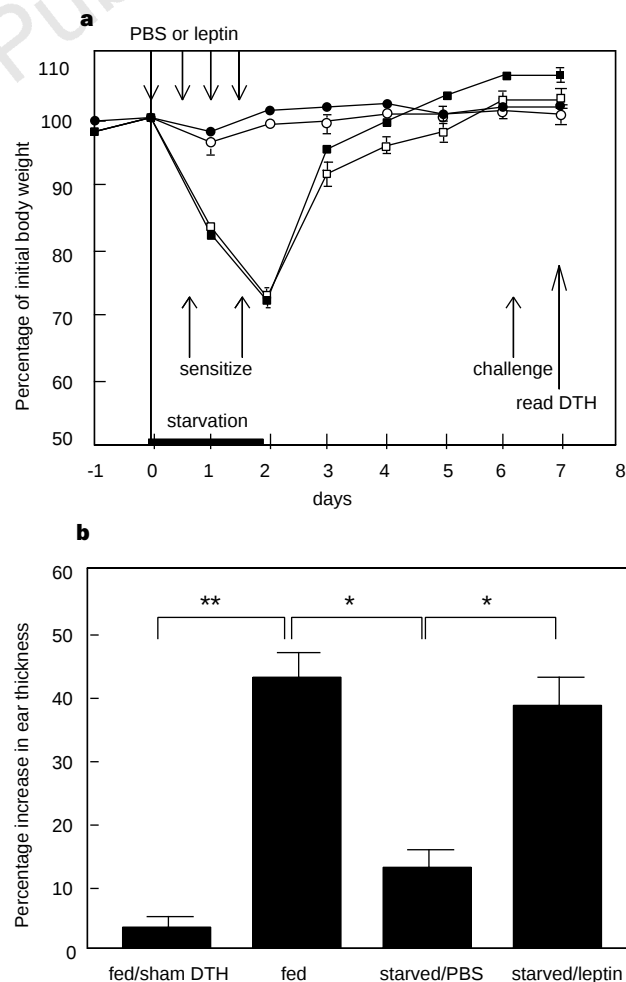


Figure 5 Leptin reverses starvation-induced immunosuppression. **a**, The effect of 48-h food deprivation on bodyweight of male C57BL/6 mice. $\bullet$ = fed/shamDTH; $\circ$ = fed; $\blacksquare$ = starved/PBS; $\square$ = starved/leptin. **b**, The effect of leptin replacement during starvation on the DTH recall response to oxazolone in male C57BL/6 mice. All data are expressed as mean percentage increase in ear thickness $\pm$ s.e.m. (n = 6 per group). * P < 0.0005, ** P < 0.00001, one-way analysis of variance with the Bonferroni adjustment for multiple group testing.

IL-2 production (not shown). *Ob/ob* responder T cells were more sensitive to exogenous leptin when compared with wild-type controls (21-fold as compared with a 5-fold increase in IFN- γ secretion; results not shown). This suggests that in the absence of leptin, Th2 responses are favoured, whereas addition of leptin seems to induce a switch to a Th1-type response in this system.

Proinflammatory cytokines such as IFN- γ can upregulate adhesion molecules²¹ and certain accessory molecules can bias a T-cell response towards the production of specific cytokines²². Morphological examination of cultures of human PBMCs and murine splenocytes showed that leptin induced marked cellular clumping at 36 hours by both human and *db/+* cells, but not by *db/db* cells (data not shown). In an MLR, leptin induced the expression of the adhesion molecules ICAM-1 (CD54) and VLA-2 (CD49b) on CD4⁺ T cells, but had no effect on the expression of other molecules that mediate adhesion (CD49a, c, d, f, CD50 or CD62L) (data not shown).

Nutritional deprivation affects immune function and also rapidly reduces the amount of circulating leptin in mice and humans^{8,9}. We investigated the capacity of exogenous leptin replacement to reverse the inhibitory effects of starvation on T-cell priming as measured by a delayed-type hypersensitivity (DTH) response. Starvation of mice for 48 hours led to a 69% reduction in the mean DTH response. This inhibition was completely reversed by injection of leptin during the period of starvation (Fig. 5).

Our observations further the hypothesis that a falling leptin concentration acts as a peripheral signal of starvation which serves to conserve energy in the face of limited reserves⁹. In this situation, energy must be preserved for vital functions such as central nervous system metabolism. Less immediately essential systems, like the reproductive axis and a finely tuned cognate immune response, which requires large-scale clonal expansion, are inhibited. Starvation is accompanied by an adaptive neuroendocrine response which includes activation of the hypothalamic–pituitary–adrenal axis and suppression of the thyroid and gonadal axes. Leptin has been shown to blunt these adaptive changes⁹. Immunoregulatory effects of these hormones are well recognized²³. The extent to which the immunomodulatory effects of leptin *in vivo* are mediated by a direct action on the CD4⁺ T cell, rather than by an indirect effect of leptin causing central suppression of the neuroendocrine response to fasting, remains to be determined. However, this regime of leptin repletion in mice only partially reverses the fasting-induced changes in corticosterone and thyroid hormones and is without effect on the starvation-induced fall in plasma insulin and glucose levels⁹. Furthermore, our *in vitro* findings indicate that leptin can modify T-cell responses directly. It is well recognized that starvation is associated with a higher frequency of infectious diseases¹⁶; our results demonstrate that leptin modulates cognate T-cell immune function and may therefore be the key link between nutritional status and an optimal immune response. □

Methods

Reagents. The human and murine recombinant leptin was purchased from R&D, Minneapolis, and from Peprotech, London. The recombinant leptin used was >97% pure, as judged by SDS–PAGE analysis and contained <0.1 ng mg⁻¹ LPS as determined by the Limulus amoebocyte lysate method. Tetanus toxoid was from Evans Medical. The appropriate mAbs were used for purifying cells for the proliferation and cytokine assays, as described²⁴.

Preparation of cells. Blood was obtained from healthy adult volunteers or from the umbilical cords of neonates (Queen Charlotte's Hospital, London). The various cell subpopulations were purified by immunomagnetic negative selection as described²⁴. The purity of the separated cell subpopulations was assessed by flow-cytometric analysis (results not shown) using a FACScan (Becton Dickinson) and was always >98%. All purified T cells were unresponsive to 72 h of culture with phytohemagglutinin (PHA) (2 μ g ml⁻¹), confirming their functional purity. Mouse responder cells from the various mouse strains used were prepared as described²⁵.

Mice. The leptin receptor mutant *db/db* mice (C57BL/Ks-db) (H-2^d), the

heterozygous *db/+* (C57BL/Ks) (H-2^d), the C57BL/6 *ob/ob* (H-2^b) and the C57BL/6 (H-2^b) mice were from Harlan Labs, Oxford, UK.

For the *in vivo* experiments, 6-week-old male C57BL/6 mice were housed in pairs under controlled temperature (21–23 °C) and a 12-h light/dark cycle (lights on at 07:00 h). Animals were handled twice daily at 09:00 and 18:00 h and food intake and body weight were recorded daily. Animals comprised four groups ($n = 6$ per group). Two groups were allowed continuous access to laboratory chow and two groups were starved for 48 h, during which they received twice-daily intraperitoneal injections of either PBS or recombinant murine leptin (1 μ g g⁻¹ initial body weight). Before and after the period of starvation, mice were allowed to feed freely. Cellular immune function was assessed by delayed-type hypersensitivity (DTH) response, which was measured essentially as described²⁶. Briefly, mice were primed on two consecutive days by application of 50 μ l of 3% oxazolone in acetone/olive oil (4:1 vol/vol) to the shaved flank. Five days later, mice were challenged with 10 μ l of 1% oxazolone applied to the ear. Ear thickness was measured at 24 h using an electronic micrometer. Mice undergoing 'sham DTH' were primed with 50 μ l of vehicle. Control challenge with 10 μ l of vehicle was performed on the contralateral ear of all mice. There was no statistical difference in the mean body weights between groups at the start of the study or at the time of challenge or reading of the DTH response.

Proliferation assays. The human mixed-lymphocyte reactions (MLRs) were done using various purified cell populations as responders (10⁵ cells per well) and the same number of allogeneic irradiated (30 Gy) PBMCs or monocytes as stimulators unless otherwise indicated, as described¹⁹. The concentration of endogenous leptin in the human AB serum used was at or below the lower detection limit in a leptin radioimmunoassay (<0.5 ng ml⁻¹; Linco, USA). Cells purified for preincubation were incubated for 12 h in the presence or absence of leptin (10 nM), washed four times, counted and plated in triplicate in 96-well plates with 10⁵ irradiated allogeneic PBMCs or monocytes for 5 days. The antigen-recall response was performed with PBMCs (10⁵ cells per well) and tetanus toxoid and measured after 5 days. Mouse MLRs were performed as described²⁵. The *ob/ob* MLRs were done in 2% autologous *ob/ob* serum.

Cytokine analysis. The production of IL-2, IL-4 and IFN- γ was measured by ELISA (Amersham Life Science) in supernatants collected after 72 h using irradiated allogeneic PBMCs or splenocytes as stimulators, in the presence or absence of leptin.

RT-PCR and Southern blotting. Total cellular RNA was extracted using RNAzol reagent (AMS Biotechnology). 5 μ g of total RNA was reverse-transcribed in a 20- μ l reaction volume using oligo(deoxythymidine)_{12–18} primer, as described²⁷. Primer selection and PCR amplification to generate a product specific for the long form (OhrB) of the human leptin receptor (nucleotide positions, 2,831–3,719) were done as described¹⁸. The nature of the expected 888-bp PCR product was verified by Southern blot analysis²⁷ using an oligonucleotide probe 5'-TATCAGATCAGCATCCCAACAT-3' (nucleotide positions 3,309–3,333 of human leptin receptor cDNA). cDNA quality was determined by the relative amplification of the human β -actin gene, which generated a 540-bp product. The T cells used in these experiments were extensively purified and were >99.5% pure on flow-cytometric staining and unresponsive to PHA. The remaining 0.5% were negative for B-cell, monocyte, NK and CD8 surface markers.

Received 20 May; accepted 26 June 1998.

- Chandra, R. K. Nutrition and immunity: Lessons from the past and new insights into the future. *Am. J. Clin. Nutr.* **53**, 1087–1101 (1991).
- Cason, J., Ainley, C. C., Wolstencroft, R. A., Norton, K. R. & Thompson, R. P. Cell-mediated immunity in anorexia nervosa. *Clin. Exp. Immunol.* **64**, 370–375 (1986).
- Polack, E. et al. Low lymphocyte interferon-gamma production and variable proliferative response in anorexia nervosa patients. *J. Clin. Immunol.* **13**, 445–451 (1993).
- Zhang, Y. et al. Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–432 (1994).
- Halaas, J. L. et al. Weight-reducing effects of the plasma protein encoded by the *obese* gene. *Science* **269**, 543–546 (1995).
- Friedman, J. M. Leptin, leptin receptors and the control of body weight. *Eur. J. Med. Res.* **2**, 7–13 (1997).
- Grinspoon, S. et al. Serum leptin levels in women with anorexia nervosa. *J. Clin. Endocrinol. Metab.* **81**, 3861–3863 (1996).
- Boden, G., Chen, X., Mozzoli, M. & Ryan, I. Effect of fasting on serum leptin in normal human subjects. *J. Clin. Endocrinol. Metab.* **81**, 3419–3423 (1996).
- Ahima, R. S. et al. Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
- Janik, J. e. et al. Interleukin 1 α increases serum leptin concentrations in humans. *J. Clin. Endocrinol. Metab.* **82**, 3084–3086 (1997).
- Sarraf, P. et al. Multiple cytokines and acute inflammation raise mouse leptin levels: potential role in inflammatory anorexia. *J. Exp. Med.* **185**, 171–175 (1997).
- Fernandes, G., Handwerker, B. S., Yunis, E. J. & Brown, D. M. Immune response in the mutant diabetic C57BL/Ks-*db/db* mouse. Discrepancies between *in vitro* and *in vivo* immunological assays. *J. Clin. Invest.* **61**, 243–250 (1978).

13. Chandra, R. K. Cell-mediated immunity in genetically obese (C57BL/6) *ob/ob* mice. *Am. J. Clin. Nutr.* **33**, 13–16 (1980).
14. Ghilardi, N. *et al.* Defective STAT signalling by the leptin receptor in diabetic mice. *Proc. Natl Acad. Sci. USA* **93**, 6231–6235 (1996).
15. Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
16. Shears, P. Epidemiology and infection in famine and disasters. *Epidemiol. Infect.* **107**, 241–251 (1991).
17. Considine, R. V. *et al.* Serum immunoreactive-leptin concentrations in normal-weight and obese humans. *N. Engl. J. Med.* **334**, 292–295 (1996).
18. Gotoda, T. *et al.* Leptin receptor gene variation and obesity: lack of association in a white British male population. *Hum. Mol. Genet.* **6**, 869–876 (1997).
19. Lombardi, G. *et al.* Are primary alloresponses truly primary? *Int. Immunol.* **2**, 9–13 (1990).
20. Abbas, A. K., Murphy, K. M. & Sher, A. Functional diversity of helper T lymphocytes. *Nature* **383**, 787–793 (1996).
21. Boehm, U., Klamp, T., Groot, M. & Howard, J. C. Cellular responses to interferon-gamma. *Annu. Rev. Immunol.* **15**, 749–795 (1997).
22. Semnani, R. T., Nutman, T. B., Hochman, P., Shaw, S. & van Seventer, G. A. Costimulation by purified intercellular adhesion molecule 1 and lymphocyte function-associated antigen 3 induces distinct proliferation, cytokine and cell surface antigen profiles in human “naïve” and “memory” CD4⁺ T cells. *J. Exp. Med.* **180**, 2125–2135 (1994).
23. Besedovsky, H. O. & del Rey, A. Immune-neuro-endocrine interactions: facts and hypotheses. *Endocr. Rev.* **17**, 64–102 (1996).
24. Young, J. L., Ramage, J. M., Hill-Gaston, J. S. & Beverley, P. C. L. *In vitro* responses of human CD45RObright and CD45RAbright T cell subsets and their relationship to memory and naïve T cells. *Eur. J. Immunol.* **27**, 2383–2389 (1997).
25. Laufer, T. M., DeKoning, J., Markowitz, J. S., Lo, D. & Glimcher, L. H. Unopposed positive selection and autoreactivity in mice expressing class II MHC only on thymic cortex. *Nature* **383**, 81–85 (1996).
26. Staite, N. D., Justen, J. M., Sly, L. M., Beaudet, A. L. & Bullard, D. C. Inhibition of delayed-type contact hypersensitivity in mice deficient in both E-selectin and P-selectin. *Blood* **88**, 2973–2979 (1996).
27. Beak, S. A. *et al.* Glucagon-like peptide-1 (GLP-1) releases thyrotropin (TSH): characterization of binding sites for GLP-1 on alpha-TSH cells. *Endocrinology* **137**, 4130–4138 (1996).

Acknowledgements. We thank C. Campagnoli, J. G. Chai, A. Di Giacomo, J. Gardiner, G. Lombardi, A. Macrae and L.-Q. Gao for technical assistance and for helpful suggestions and D. Gray, H. Reiser and D. Smith for critically reviewing the manuscript. G.M.L., J.K.H. and R.J.B. are MRC Clinical Training Fellows.

Correspondence and requests for materials should be addressed to R.L.L. (e-mail: riechler@rmps.ac.uk).

Csk controls antigen receptor-mediated development and selection of T-lineage cells

Christian Schmedt[§], Kaoru Saijo[§], Tetsuhiro Niidome[†], Ralf Kühn[‡], Shinichi Aizawa[†] & Alexander Tarakhovskiy^{*}

^{*}Laboratory for Lymphocyte Signalling and [‡]Department of Immunology, Institute for Genetics, University of Cologne, Weyertal 121, 50931 Cologne, Germany

[†]Department of Morphogenesis, Institute of Molecular Embryology and Genetics, Kumamoto University School of Medicine, Kumamoto-860, Japan

[§]These authors contributed equally to this work

The development and function of $\alpha\beta$ T lymphocytes depend on signals derived from pre-T and $\alpha\beta$ T cell receptors (preTCR and $\alpha\beta$ TCR) (reviewed in refs 1, 2). The engagement of these receptors leads to the activation of Lck and Fyn^{3,4}, which are protein tyrosine kinases (PTKs) of the Src family. It remains unclear to what extent the activation of Src-family PTKs can direct the differentiation steps triggered by preTCR and $\alpha\beta$ TCR. Here we show that the inactivation of the negative regulator of Src-family PTKs, carboxy-terminal Src kinase (Csk)⁵, in immature thymocytes abrogates the requirement for preTCR, $\alpha\beta$ TCR and major histocompatibility complex (MHC) class II for the development of CD4⁺CD8⁺ double-positive and CD4⁺ single-positive thymocytes as well as peripheral CD4 $\alpha\beta$ T-lineage cells. These data show that Csk and its substrates are required to establish preTCR/ $\alpha\beta$ TCR-mediated control over the development of $\alpha\beta$ T cells.

Signals transmitted through the preTCR are required for the generation of CD4⁺CD8⁺ double-positive (DP) thymocytes¹. The binding of MHC class I/II plus peptide to $\alpha\beta$ TCR and CD4/CD8 co-receptors is essential for the differentiation of DP cells into CD4⁺ or CD8⁺ single-positive (SP) thymocytes and peripheral T cells

(reviewed in ref. 6). The early block in thymocyte development in Lck^{−/−}; Fyn^{−/−} mice^{7,8}, and the ability of constitutively active Lck to trigger the preTCR-independent generation of DP cells⁹, shows that preTCR-mediated thymocyte development depends on Src-family PTKs. However, the failure of constitutively active Lck to induce $\alpha\beta$ TCR-independent progression past the DP stage of thymocyte development suggests that differentiation into SP thymocytes and peripheral T cells may require the concerted activation of Lck and Fyn, or may depend on additional factors. The activity of Src-family PTKs is negatively regulated by Csk and is increased in Csk-deficient cells^{10,11}. We hypothesized that the inactivation of Csk and concomitant activation of Src-family PTKs in double-negative (DN) thymocytes may result in the preTCR/ $\alpha\beta$ TCR-independent development of mature $\alpha\beta$ T-lineage cells. We inactivated the gene encoding Csk conditionally, because mice homozygous for a *csk* null mutation die before the onset of embryonic lymphopoiesis^{10,11}.

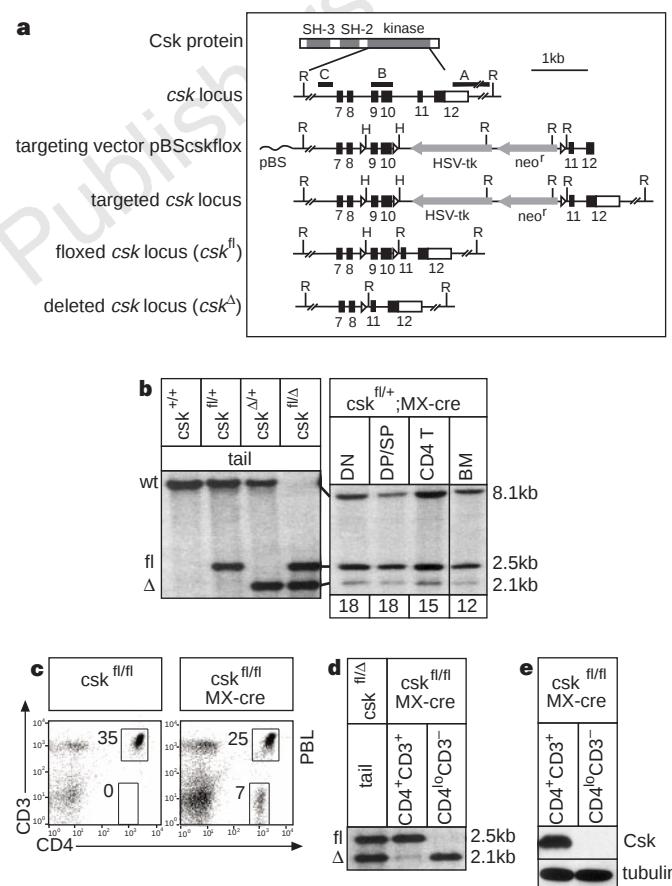


Figure 1 Conditional inactivation of *csk* results in the appearance of CD3⁺CD4^{lo} peripheral cells. **a**, Domain structure of Csk protein, partial structure of the *csk* locus, and maps of the targeting vector and the targeted *csk* locus before and after Cre-mediated recombinations. Numbered rectangles represent exons, coding regions are filled; open triangles represent *loxP* sites; grey arrows indicate the position and transcriptional orientation of selection marker genes. Restriction sites are: R, *EcoRI*; H, *HindIII*. Probes A, B and C used in Southern blots are shown as bars. **b**, Uninduced deletion of the floxed *csk* fragment in cells of *csk*^{fl/+}; MX-cre and control mice. The size of the DNA fragment corresponding to wild-type (wt), floxed (fl) and deleted (Δ) *csk* alleles and per cent of deletion are indicated. **c**, CD3⁺CD4^{lo} cells in the peripheral blood of *csk*^{fl/fl}; MX-cre mice. PBLs of *csk*^{fl/fl} and *csk*^{fl/fl}; MX-cre mice were stained with anti-CD3 and anti-CD4 antibodies. Numbers in larger font show percentages of gate cells. **d**, Deletion of *csk* in splenic cells. Genomic DNA of purified CD3⁺CD4⁺ and CD3⁺CD4^{lo} cells or tail of *csk*^{fl/fl}; MX-cre or *csk*^{fl/Δ} mice, respectively, was analysed by Southern blotting. **e**, Csk protein expression in CD3⁺CD4⁺ and CD3⁺CD4^{lo} cells. Lysates of purified cells were analysed by western blotting with anti-Csk antibody, and reprobed with an anti- α -tubulin antibody.

^{||}Present address: CNS Research Unit, Department of Drug Discovery, Tsukuba Research Laboratories, Eisai Co.Ltd, 1–3 Tokodai 5-chome, Tsukuba-shi, Ibaraki 300-26, Japan.

To conditionally inactivate the *csk* gene, exons 9 and 10 (which encode essential parts of the kinase domain of Csk^{5,12}) were flanked with *loxP* sites (floxed) by homologous recombination in embryonic stem (ES) cells (Fig. 1a), and mice homozygous for the floxed allele of *csk* were generated (*csk*^{fl/fl}). To inactivate *csk* in various cell types, *csk*^{fl/fl} mice were crossed with MX-cre mice, which carry a Cre-recombinase transgene under the control of the type I interferon-inducible MX promoter¹³. However, analysis of heterozygous *csk*^{fl/+}; MX-cre mice showed uninduced deletion of the floxed allele of *csk* in up to 18% of thymocytes, CD4 T cells and bone-marrow cells, as quantified by phosphorimage analysis (Fig. 1b). It was therefore not necessary to induce further *csk* inactivation with interferon.

The analysis of peripheral lymphocytes of *csk*^{fl/fl}; MX-cre mice reveals a population of surface CD4-positive but CD3^ε- and αβTCR-negative (CD3⁺, αβTCR⁻) cells (Fig. 1c and data not shown). These cells make up $3.2 \pm 2.5\%$ ($n = 33$) of the peripheral blood lymphocyte (PBL) population of *csk*^{fl/fl}; MX-cre mice (Fig. 1c) and are Csk-deficient (Fig. 1d, e). CD4 surface expression on αβTCR⁻CD3⁺ cells is reduced two- to threefold (Fig. 1c). However, similar levels of CD4 messenger RNA in control and CD4^{lo} cells (data not shown) confirm the CD4-lineage identity of CD4^{lo} cells, and suggest that Csk is involved in regulating CD4 surface expression. The presence of rearranged T-cell antigen receptor (TCR)-α genes in αβTCR⁻CD3⁺CD4^{lo} cells shows that they belong to the αβT cell lineage (Fig. 2b, left). Furthermore, like wild-type αβT

cells, they express high levels of Thy-1 and CD5, but do not express γδTCR or the natural killer cell marker DX5 (ref. 14; data not shown). Peripheral αβTCR⁻CD3⁺CD4^{lo} T-lineage cells are likely to derive from the thymus, as cells of similar phenotype are present among DP and CD4 SP thymocytes (Fig. 2a, centre).

Because the deletion of *csk* in *csk*^{fl/+}; MX-cre mice occurs in a similar fraction of DN or DP/SP thymocytes and peripheral CD4 T cells (Fig. 1b), the appearance of αβTCR⁻CD3⁺CD4^{lo} cells could result from the loss of *csk* at any of these developmental stages. However, it is plausible that αβTCR⁻CD3⁺CD4^{lo} cells develop from DN cells in which the deletion of *csk* and the concurrent activation of Src-family PTKs occur before productive TCRβ gene rearrangement. To show that Csk-deficient αβTCR⁻DN thymocytes can develop into αβTCR⁻CD3⁺CD4^{lo} thymocytes and peripheral T lineage cells independent of preTCR and αβTCR, *csk*^{fl/fl}; MX-cre mice were crossed with TCRβ^{-/-} mice¹⁵. Thymi of TCRβ^{-/-} mice are small ($1-2 \times 10^7$ cells) and contain largely DN cells and γδ T-lineage-related DP and SP cells¹⁵. In contrast, thymi of TCRβ^{-/-}; *csk*^{fl/fl}; MX-cre mice are large ($9.6 \times 10^7 \pm 1.5 \times 10^7$ cells, $n = 7$) and contain both DP and CD4 SP cells at proportions close to those in control mice (Fig. 2a). These cells are Csk-deficient, surface αβTCR⁻CD3⁺CD4^{lo}DX5⁻, do not express γδTCR (Fig. 2a) and, according to the surface expression levels of the maturation markers HSA (ref. 16) and CD69 (ref. 17), are as mature as cells of the corresponding subpopulations of control thymocytes (data not shown). CD8 SP cells are largely absent from the thymi and lymph nodes of TCRβ^{-/-}; *csk*^{fl/fl}; MX-cre mice (Fig. 2a, right). This suggests that, in the absence of αβTCR, development of CD8 cells cannot be mimicked by the activation of Src-family PTKs, or that Csk has a positive role in controlling commitment to the CD8 lineage. Csk-deficient thymocytes in TCRβ^{-/-}; *csk*^{fl/fl}; MX-cre mice can develop further into peripheral αβ-lineage cells, which make up $23 \pm 8\%$ ($n = 24$) of the PBL population and constitute up to 40% of cells in mesenteric lymph nodes. These cells contain rearranged TCRα genes and have other features of mature, naive peripheral T cells¹⁸ (Thy-1⁺, CD5⁺, CD25⁻, CD69⁻, CD62L^{hi}, CD44^{lo}) (Fig. 2b and data not shown).

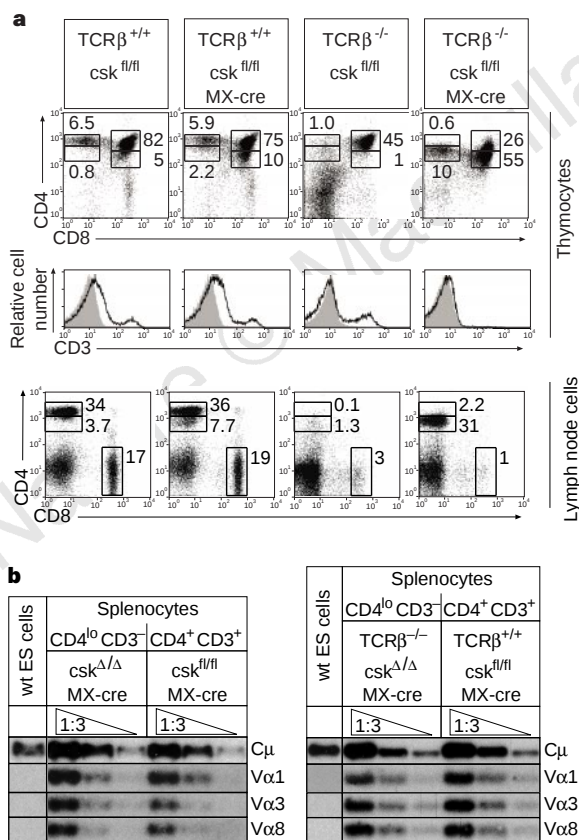


Figure 2 αβTCR-independent development of Csk-deficient T-lineage cells. **a**, Expression of surface markers on thymocytes (top) and lymph-node cells (bottom) of TCRβ^{-/-}; *csk*^{fl/fl}; MX-cre and control mice. Numbers in larger font indicate percentages of gated cells. Histograms show the expression levels of CD3 on total thymocytes (lines) or negative staining controls (shading). **b**, Rearrangements of TCRα genes in CD4^{lo}CD3⁺ *csk*-deficient cells of αβTCR^{+/+} (left) and αβTCR^{-/-} (right) mice. Rearranged TCRα genes were identified by PCR using genomic template DNA at three different dilutions from indicated cell populations. Primer pairs specific for the constant region of the Igμ gene²⁷ and TCRα rearrangements involving Vα1, Vα3 or Vα8 and JaTT11 (ref. 26) were used.

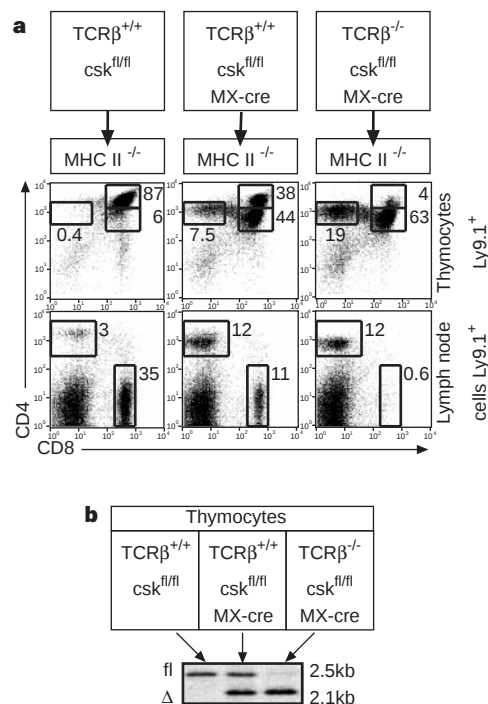


Figure 3 MHC class II- and αβTCR-independent development of Csk-deficient T-lineage cells. **a**, Subpopulations of donor-derived (Ly9.1⁺) thymocytes and lymph-node cells are shown. Numbers in larger font indicate the percentages of gated cells. **b**, Southern blot analysis, as described for Fig. 1b, of total thymocyte DNA of reconstituted MHC class II-deficient mice.

The cytoplasmic domain of CD4 binds Lck¹⁹, and interaction of MHC class II molecules with CD4 can lead to preTCR-independent development of DP thymocytes²⁰. To exclude the possibility that interactions between MHC class II and CD4 cause the appearance of $\alpha\beta$ TCR⁺CD3⁺CD4^{lo} Csk-deficient cells, we examined whether these cells develop in the absence of MHC class II. Bone-marrow cells of control or TCR $\beta^{-/-}$; *csk*^{fl/fl}; MX-cre mice were transferred into lethally irradiated MHC class II-deficient mice²¹, and we analysed the development of donor-derived T cells. Consistent with the previously described arrest of CD4 T-lineage development in MHC class II-deficient mice^{21–23}, *csk*^{fl/fl} CD4 T cells cannot develop in the absence of MHC class II (Fig. 3a, left). Bone-marrow cells of *csk*^{fl/fl}; MX-cre mice give rise to DP thymocytes expressing either high or low levels of CD4. However, only CD4^{lo} SP thymocytes and CD4^{lo} peripheral T-lineage cells can develop from these DP thymocytes in MHC class II-deficient mice (Fig. 3a, centre). Finally, the transfer of TCR $\beta^{-/-}$; *csk*^{fl/fl}; MX-cre bone-marrow cells results in the generation exclusively of Csk-deficient donor-derived thymocytes (Fig. 3b). DP thymocytes show uniformly low CD4 expression and differentiate into CD4^{lo} SP and peripheral CD4^{lo} T-lineage cells (Fig. 3a, right).

Although $\alpha\beta$ TCR⁺CD3⁺CD4^{lo} Csk-deficient T-lineage cells are a

priori unable to fulfil normal T-cell functions *in vivo*, we confirmed their $\alpha\beta$ T-lineage identity by using functional assays. Purified Csk-deficient T-lineage cells stimulated with phorbol-12-myristate-13-acetate (PMA) plus Ca²⁺-ionophore (A23187) change the expression of T-cell activation markers in a manner similar to control cells (Fig. 4a). Csk-deficient cells proliferate and produce interleukin (IL)-2 (29 U ml⁻¹ versus 52 U ml⁻¹ from control cells) in response to PMA plus A23187, but not PMA plus anti-CD3 ϵ (Fig. 4b). When differentiated *in vitro* in the presence of IL-12 after stimulation with PMA plus A23187, Csk-deficient cells assume a Th1-like phenotype, as defined by secretion of interferon (IFN)- γ (Fig. 4c).

These data show that, without Csk, the development and maintenance of $\alpha\beta$ T cells, as well as their commitment to the CD4 lineage, are independent of preTCR, $\alpha\beta$ TCR and MHC class II. Increased catalytic activity for Lck (two- to threefold) and Fyn (four- to fivefold) is observed in Csk-deficient thymocytes (Fig. 4d). Hyperactivity of Src-family PTKs appears sufficient to allow preTCR/ $\alpha\beta$ TCR-independent differentiation of $\alpha\beta$ T-lineage cells. In normal $\alpha\beta$ T-cell development, however, Csk prevents unregulated differentiation by inhibiting Src-family PTKs. This inhibition is overruled only by appropriate preTCR/ $\alpha\beta$ TCR-mediated differentiation signals. Thus, the negative control that Csk exerts over Src-family PTKs is a prerequisite for antigen receptor-mediated control of $\alpha\beta$ T-cell development. □

Methods

Generation of *csk*^{fl/fl} mice. To create the targeting vector pBSckflo, a single loxP site was introduced into a Spel site in intron 8 and an HSVtk-neo^r selection marker cassette flanked by loxP sites²⁴ was inserted into an XbaI site in intron 10 of *csk* (Fig. 1). ES cells at embryonic day 14.1 were transfected and selected as described²⁴. Homologous recombinants were identified by Southern blot analysis (EcoRI digest, probe A; HindIII digest probe B). The floxed HSVtk-neo^r cassette was removed from targeted ES cells by transient transfection with a Cre expression vector as described²⁴. Cells of floxed or deleted clones identified by Southern blot analysis (EcoRI digest, probe C) were used to generate mice carrying floxed (*csk*^{fl}) or deleted (*csk*^Δ) *csk* alleles.

Mouse strains. Mice were maintained in a conventional animal facility and analysed at 6–16 weeks of age. The *csk*^{fl/fl}; MX-cre, *csk*^{fl/fl}; *csk*^{fl/fl}; MX-cre and TCR $\beta^{-/-}$; *csk*^{fl/fl}; MX-cre mice have a mixed genetic background of 129, CB.20 and C57B1/6. MHC class II-deficient mice (α A^{-/-}), TCR $\beta^{-/-}$ and DO11.10 TCR transgenic mice have C57B1/6, 129/B6 mixed and Balb/c genetic backgrounds, respectively.

Flow cytometry and purification of cells. Monoclonal antibodies against the following antigens were purchased from Pharmingen (San Diego, CA): CD3 ϵ (145-2C11), CD4 (RM4-5), CD8 (53-6.7), CD24/HSA (M1/69), CD25 (7D4), CD44 (IM7), CD62L (Me114), CD69 (H1.2F3), natural killer cells (DX5), Ly9.1 (30C7), TCR β (H57-597), TCR δ (GL3), murine IL-4 (11B11) and murine IFN- γ (XMG1.2). The following monoclonal antibodies were prepared: RA3-6B2 (anti-B220), CFO-1 (anti-CD90/Thy1) and M1/70 (anti-Mac-1). Cells were purified from single-cell suspensions by magnetic cell sorting (MACS) using depletion or enrichment protocols according to manufacturer's instructions (Miltenyi Biotec, Bergisch Gladbach, Germany). The following antibody combinations and protocols were used for the purification of cells. DN thymocytes: anti-CD4, anti-CD8, anti-B220 and anti-Mac-1 depletion; DP/SP thymocytes: anti-CD4 and anti-CD8 enrichment; CD4 T cells: anti-CD4 enrichment. To yield CD4⁺CD3⁺ and CD4^{lo}CD3⁺ cells from *csk*^{fl/fl}; MX-cre mice, MACS-sorted CD4 cells were further purified by fluorescence-activated cell sorting (FACS) on a FACstar (Becton Dickinson). MACS-sorted cells were more than 95% pure; FACS-sorted cells were more than 98% pure.

Bone-marrow transfer. T-cell-depleted bone-marrow cells derived from mice expressing the Ly9.1 surface marker²⁵ were transferred intravenously (10⁷ cells per mouse) into lethally irradiated (950 rad) Ly9.1-negative mice. The recipient mice were killed 7–8 weeks after the transfer, and donor-derived surface Ly9.1⁺ thymocyte and lymph-node cells were analysed by FACS.

TCR α gene rearrangements. TCR α gene rearrangements were analysed by the polymerase chain reaction (PCR) followed by Southern blot hybridization, as described^{26,27}.

Proliferation assay and *in vitro* differentiation. Proliferation assays were as

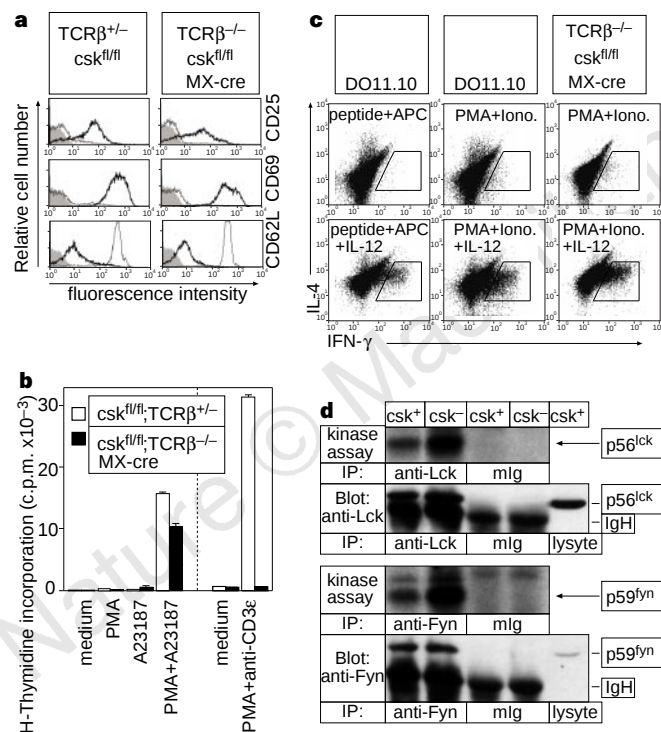


Figure 4 Functional properties of Csk-deficient $\alpha\beta$ T-lineage cells. **a**, T-cell activation-marker expression after stimulation *in vitro*. CD3⁺CD4⁺ and CD3⁺CD4^{lo} cells of TCR $\beta^{+/+}$; *csk*^{fl/fl} or TCR $\beta^{-/-}$; *csk*^{fl/fl}; MX-cre mice, respectively, were incubated for 24 h in medium (thin lines) or with PMA plus the Ca²⁺-ionophore A23187 (thick lines). Shaded histograms show negative staining controls. **b**, Proliferative response of CD3⁺CD4⁺ and CD3⁺CD4^{lo} cells to stimulation with PMA plus A23187, PMA plus anti-CD3 ϵ monoclonal antibody or controls. **c**, Expression of IFN- γ in CD4 cells after *in vitro* differentiation. CD4 cells of ovalbumin-specific $\alpha\beta$ TCR transgenic (DO11.10) or TCR $\beta^{-/-}$; *csk*^{fl/fl}; MX-cre mice were differentiated *in vitro* in the presence or absence of IL-12 in combination with antigenic peptide (OVA_{323–339}) plus antigen-presenting cells, or PMA plus A23187, as indicated. IFN- γ -producing cells are framed. **d**, Autophosphorylation of Lck and Fyn in Csk-deficient thymocytes. Fyn and Lck were precipitated from Csk-positive or Csk-negative thymocyte lysates of control TCR $\beta^{+/+}$; *csk*^{fl/fl} or TCR $\beta^{-/-}$; *csk*^{fl/fl}; MX-cre mice, respectively. Immune complexes were used for *in vitro* kinase assay or western blotting. Specificity of precipitating or blotting antibodies was controlled by mouse immunoglobulin (mlg) immunoprecipitate or whole-cell lysate of Csk-positive thymocytes, respectively.

described²⁸. IL-2 production was measured by enzyme-linked immunosorbent assay (ELISA) with the monoclonal antibodies JES6-1A12 and JES6-5H4 (Pharmingen). Purified naive splenic CD4 cells were differentiated into Th1 cells as described²⁹ or with PMA plus A23187. Briefly, 10^6 CD4 cells were stimulated with 5 ng ml^{-1} PMA (Calbiochem), 100 ng ml^{-1} A23187 (Calbiochem) and 1.5 ng ml^{-1} IL-2 (R&D, Wiesbaden). After 4 days of culture, cells were restimulated with PMA plus A23187 and analysed for intracellular cytokine expression²⁹.

Western blot analysis and *in vitro* kinase assays. Cells were lysed in 1% NP-40 lysis buffer (10 mM Tris/HCl, pH 7.8, 150 mM NaCl, 4 mM EDTA, 10% glycerol, 2 mM Na_3VO_4 , 100 mM NaF, $2 \mu\text{g ml}^{-1}$ aprotinin, $1 \mu\text{g ml}^{-1}$ leupeptin, 1 mM PMSF). Lysates were resolved by 10% SDS-PAGE and transferred to PVDF membrane (Immobilon-P, Millipore). Membranes were incubated with anti-Csk (Santa Cruz), anti- α -tubulin (DM1A, Sigma), anti-Lck (CT, UBI, Lake Placid) or anti-Fyn (FYN3, Santa Cruz) antibodies and proteins were detected by chemiluminescence (SuperSignal, Pierce). For *in vitro* kinase assays, NP-40 lysates of 5×10^7 thymocytes were precipitated with anti-Lck (3A5, UBI) or anti-Fyn (FYN15, Santa Cruz) antibodies and protein-A sepharose. Immune complexes were incubated in kinase buffer (10 mM HEPES, pH 7.4, 5 mM MgCl_2 , 5 mM MnCl_2) with $15 \mu\text{Ci}[\gamma\text{-}^{32}\text{P}]\text{ATP}$ for 2 min at 37°C and resolved by 10% SDS-PAGE. The conditions of the kinase assay were tested to fall within the linear range of the kinase reaction. Autophosphorylation of Lck and Fyn was quantified by phosphorimage analysis of dried gels with Fuji Bas1000.

Received 5 June; accepted 8 July 1998.

1. von Boehmer, H. & Fehling, H. J. Structure and function of the pre-T cell receptor. *Annu. Rev. Immunol.* **15**, 432–452 (1997).
2. Mombaerts, P. Lymphocyte development and function in T-cell receptor and RAG-1 mutant mice. *Int. Rev. Immunol.* **13**, 43–63 (1995).
3. Alberola-Ila, J., Takaki, S., Kerner, J. D. & Perlmutter, R. M. Differential signaling by lymphocyte antigen receptors. *Annu. Rev. Immunol.* **15**, 125–154 (1997).
4. Weiss, A. & Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **76**, 263–274 (1994).
5. Nada, S., Okada, M., MacAuley, A., Cooper, J. A. & Nakagawa, H. Cloning of a complementary DNA for a protein-tyrosine kinase that specifically phosphorylates a negative regulatory site of p60^{src} . *Nature* **351**, 69–72 (1991).
6. Jameson, S. C., Hogquist, K. A. & Bevan, M. J. Positive selection of thymocytes. *Annu. Rev. Immunol.* **13**, 93–126 (1995).
7. Groves, T. *et al.* Fyn can partially substitute for Lck in T lymphocyte development. *Immunity* **5**, 417–428 (1996).
8. van Oers, N. S., Lowin-Kropf, B., Finlay, D., Connolly, K. & Weiss, A. $\alpha\beta$ T cell development is abolished in mice lacking both Lck and Fyn protein tyrosine kinases. *Immunity* **5**, 429–436 (1996).
9. Mombaerts, P., Anderson, S. J., Perlmutter, R. M., Mak, T. W. & Tonegawa, S. An activated lck transgene promotes thymocyte development in RAG-1 mutant mice. *Immunity* **1**, 261–267 (1994).
10. Nada, S. *et al.* Constitutive activation of Src family kinases in mouse embryos that lack Csk. *Cell* **73**, 1125–1135 (1993).
11. Imamoto, A. & Soriano, P. Disruption of the csk gene, encoding a negative regulator of Src family tyrosine kinases, leads to neural tube defects and embryonic lethality in mice. *Cell* **73**, 1117–1124 (1993).
12. Hanks, S. K. & Hunter, T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576–596 (1995).
13. Kühn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
14. Moore, T. A., von Freeden-Jeffry, U., Murray, R. & Zlotnik, A. Inhibition of gamma delta T cell development and early thymocyte maturation in IL-7^{-/-} mice. *J. Immunol.* **157**, 2366–2373 (1996).
15. Mombaerts, P. *et al.* Mutations in T-cell antigen receptor genes α and β block thymocyte development at different stages. *Nature* **360**, 225–231 (1992).
16. Wilson, A., Day, L. M., Scollay, R. & Shortman, K. Subpopulations of mature murine thymocytes: properties of CD4⁺CD8⁺ and CD4⁺CD8⁻ thymocytes lacking the heat-stable antigen. *Cell Immunol.* **117**, 312–326 (1988).
17. Swat, W., Dessing, M., von Boehmer, H. & Kisielow, P. CD69 expression during selection and maturation of CD4⁺8⁺ thymocytes. *Eur. J. Immunol.* **23**, 739–746 (1993).
18. Sprent, J. & Tough, D. F. Lymphocyte life-span and memory. *Science* **265**, 1395–1400 (1994).
19. Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56^{lck} . *Cell* **55**, 301–308 (1988).
20. Normant, A. M., Forbush, K. A., Nguyen, N., Malissen, M. & Perlmutter, R. M. Replacement of pre-T cell receptor signaling functions by the CD4 coreceptor. *J. Exp. Med.* **185**, 121–130 (1997).
21. Köntgen, F., Süss, G., Stewart, C., Steinmetz, M. & Bluethmann, H. Targeted disruption of the MHC class II α gene in C57BL/6 mice. *Int. Immunol.* **5**, 957–964 (1993).
22. Grusby, M. J., Johnson, R. S., Papaioannou, V. E. & Glimcher, L. H. Depletion of CD4⁺ T cells in major histocompatibility complex class II-deficient mice. *Science* **253**, 1417–1420 (1991).
23. Grogrove, D. *et al.* Mice lacking MHC class II molecules. *Cell* **66**, 1051–1066 (1991).
24. Torres, R. & Kühn, R. *Laboratory Protocols for Conditional Gene Targeting* (Oxford Univ. Press, 1997).
25. Ledbetter, J. A. & Herzenberg, L. A. Xenogeneic monoclonal antibodies to mouse lymphoid differentiation antigens. *Immunol. Rev.* **47**, 63–90 (1979).
26. Anderson, S. J., Abraham, K. M., Nakayama, T., Singer, A. & Perlmutter, R. M. Inhibition of T-cell receptor β -chain gene rearrangement by overexpression of the non-receptor protein tyrosine kinase p56^{lck} . *EMBO J.* **11**, 4877–4886 (1992).
27. Krotkova, A., von Boehmer, H. & Fehling, H. J. Allelic exclusion in pT α -deficient mice: no evidence for cell surface expression of two T cell receptor (TCR)- β chains but less efficient inhibition of endogenous V β $\rightarrow$ (D) β rearrangements in the presence of a functional TCR- β transgene. *J. Exp. Med.* **186**, 767–775 (1997).
28. Saijo, K., Park, S. Y., Ishida, Y., Arase, H. & Saito, T. Crucial role of Jak3 in negative selection of self-reactive T cells. *J. Exp. Med.* **183**, 351–356 (1997).
29. Murphy, E. *et al.* Reversibility of T helper 1 and 2 populations is lost after long-term stimulation. *J. Exp. Med.* **183**, 901–913 (1996).

Acknowledgements. We thank A. Berns and H. Jacobs for TCR β ^{-/-} mice; H. Bluethmann and I. Förster for $\text{A}\alpha$ ^{-/-} mice; D. Loh and M. Löhning for DO11.10 mice; C. Göttinger, B. Meiners, S. Irlenbusch and C. Uthoff-Hachenberg for technical assistance; and J. Howard and K. Rajewsky for discussions and critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 243. C.S. was supported by the Fonds der Chemie. K.S. was supported by an EMBO long-term fellowship.

Correspondence and requests for material should be addressed to C.S. (e-mail: c.schmedt@uni-koeln.de) and A.T. (e-mail: sasha@mac.genetik.uni-koeln.de).

FGF-mediated mesoderm induction involves the Src-family kinase Laloo

Daniel C. Weinstein, Jennifer Marden, Francesca Carnevali* & Ali Hemmati-Brivanlou

Department of Molecular Vertebrate Embryology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

During embryogenesis, inductive interactions underlie the development of much of the body plan. In *Xenopus laevis*, factors secreted from the vegetal pole induce mesoderm in the adjacent marginal zone; members of both the transforming growth factor- β (TGF- β) and fibroblast growth factor (FGF) ligand families seem to have critical roles in this process¹. Here we report the identification and characterization of *laloo*, a novel participant in the signal transduction cascade linking extracellular, mesoderm-inducing signals to the nucleus, where alteration of cell fate is

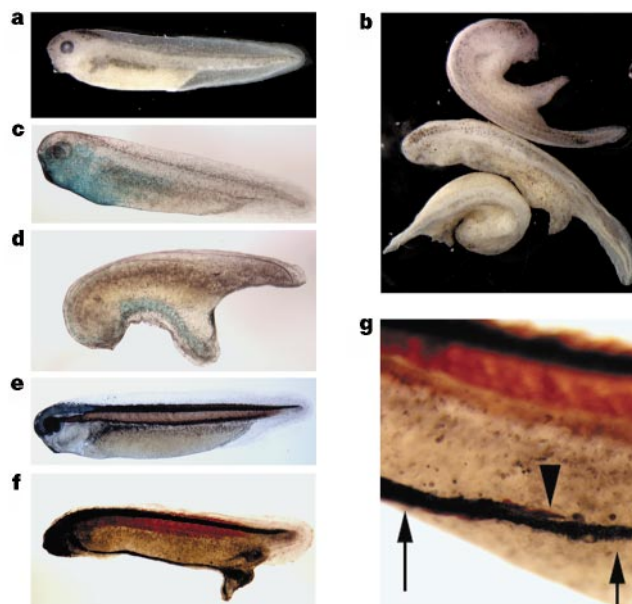


Figure 1 Effects of injection of *laloo* into embryos. Embryos were injected with 100 pg β -galactosidase RNA (**c**, **d**) and/or 750 pg (**b**, **d**, **f**, **g**) 27A1JA (*laloo*) RNA. All embryos are stage 33–35. **a**, Uninjected embryo. **b**, Dorsal view of embryos injected with 27A1JA. **c**, Embryo injected with β -galactosidase (β -Gal) RNA and stained with 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) as a substrate. **d**, Embryo co-injected with 27A1JA and β -Gal RNA, and stained with X-gal as a substrate. **e–g**, Lateral views of embryos probed with neural-specific (blue stain) (6F11 (ref. 27)) and somite-specific (red stain) (12/101 (ref. 28)) antibodies. **e**, Uninjected embryo. **f**, **g**, Embryo injected with 27A1JA RNA. Panel **g** is a detail of the trunk of embryo in **f**; arrows indicate ectopic neural tissue, arrowhead indicates ectopic mesoderm. Embryos in **c–g** were cleared in 2:1 benzyl benzoate/benzyl alcohol.

* Present address: Centro di Studio per gli Acidi Nucleici, CNR, Università di Roma La Sapienza, P. le Aldo Moro n 5, 00185 Roma, Italy.

driven by changes in gene expression. Overexpression of *laloo*, a member of the Src-related gene family, in *Xenopus* embryos gives rise to ectopic posterior structures that frequently contain axial tissue. *Laloo* induces mesoderm in *Xenopus* ectodermal explants; this induction is blocked by reagents that disrupt the FGF signalling pathway. Conversely, expression of a dominant-inhibitory *Laloo* mutant blocks mesoderm induction by FGF and causes severe posterior truncations *in vivo*. This work provides the first evidence that a Src-related kinase is involved in vertebrate mesoderm induction.

In an attempt to isolate factors involved in patterning of the body axis, we constructed and screened a *Xenopus* gastrula expression library. Early cleavage stage embryos were injected in the animal pole with RNA from fractionated library pools. One clone, 27AIJA, generated ectopic structures that resemble tails in over 90% of injected embryos ($n = 102$) (Fig. 1a, b). In addition, a reduction of anterior structures, including eyes, was often observed (Fig. 1, and not shown). Co-injection of 27AIJA and β -Gal RNA revealed that injected cells contribute directly to ectopic structures (Fig. 1c, d). Whole-mount immunohistochemistry demonstrated that these ectopic structures often contain neural tissue (17 out of 32 embryos) and paraxial mesoderm (5 out of 32 embryos) (Fig. 1e–g). Thus, overexpression of 27AIJA generates ectopic, posterior structures that frequently contain axial and paraxial tissue.

The 27AIJA cDNA contains an open reading frame of 496 amino

acids (Fig. 2a); this sequence shows homology with the Src family of intracellular tyrosine kinases², and includes putative SH3, SH2 and kinase domains (Fig. 2b). Several lines of evidence indicate that 27AIJA encodes a novel member of this gene family. First, the unique, amino-terminal domain of 27AIJA resembles that of two amniote family members, Lyn and Hck, almost equally (26% and 31% similarity, respectively)². Second, the complete coding sequence of 27AIJA is more than twice as divergent from its closest amniote relative (Hck, 38–40%) than are the sequences of other cloned *Xenopus* Src-family genes (*Xsrc*, *Xyes*, *Xfyn*, *Xlyn*) from their amniote homologues (3–18% (ref. 2; *Xlyn* sequence ID no. 2114076)). Finally, 27AIJA is less closely related to amniote Hck (38–40% divergence) than is the putative *Xenopus* homologue of the related Lyn gene (31–34% divergence). Thus, 27AIJA encodes a novel factor that can induce posterior, ectopic axes; we have named this gene *laloo*, after a nineteenth-century circus performer who had a small, headless twin protruding from his breastbone.

To better define the embryonic function of *laloo*, we tested its activity in ectodermal explant (animal cap) assays. At midgastrula stages, reverse transcription in conjunction with polymerase chain reaction (RT-PCR) revealed that *laloo*-injected caps express both *Xbra*, a pan-mesodermal marker, and *Xwnt8*, a marker of ventro-lateral mesoderm^{3,4}, but not *chordin*, a dorsal mesodermal marker⁵ (Fig. 3a, lanes 1–4). At late neurula stages, *laloo*-expressing caps show strong expression of *HoxB9* (*XlHbox6*), which at this stage is

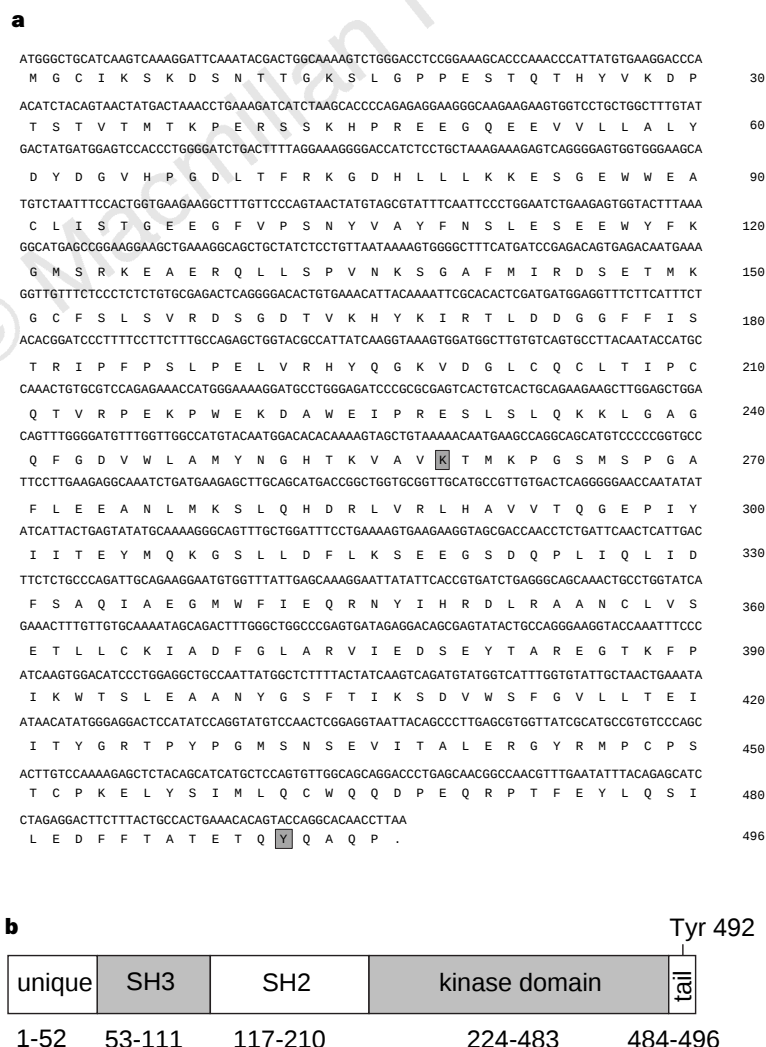


Figure 2 *laloo* is a novel Src-related gene. **a**, Nucleotide sequence and predicted open reading frame of *laloo*. Residues mutated in this study (K259, Y492) are boxed. **b**, Schematic of the *Laloo* protein tyrosine kinase. The *Xenopus* *Laloo*

protein is drawn approximately to scale, showing the positions of domains conserved throughout the Src gene family.

expressed in both lateral mesoderm and the spinal cord⁶ (Fig. 3b, lanes 1–4). Because NCAM, a pan-neural marker⁷, is not induced, we conclude that the *HoxB9* expression induced by Laloo at this stage represents mesodermal tissue. At high doses, Laloo induces the expression of *muscle actin*, a marker of paraxial mesoderm⁸ (Fig. 3b; lane 1); this result demonstrates that high levels of Laloo expression give rise to more dorsal fates than do lower doses. Figure 3c shows that Laloo also induces mesodermal markers in the context of the whole embryo at gastrula stages. Whereas control embryos express *Xbra* in a ring at the marginal zone at midgastrula stages, animal

pole injections of *laloo* RNA give rise to ectopic *Xbra* expression in the animal pole (39%, $n = 18$; Fig. 3c). These data indicate that overexpression of Laloo induces mesoderm, both in the animal cap assay and *in vivo*.

Mesoderm induction in *Xenopus* occurs between cleavage and early gastrula stages¹; *laloo* is expressed throughout this period (Fig. 3d). *laloo* RNA is present maternally; this early expression is, however, greatly diminished by late gastrula stages. Zygotic *laloo* expression initiates after late neurula stages. Using a combination of whole-mount *in situ* hybridization and microdissection techniques, we observed no tissue-specific localization of *laloo* RNA through neurula stages (not shown). The ubiquitous, early expression of this gene is consistent with a role for *laloo* in endogenous mesoderm induction.

Members of both the TGF- β and FGF ligand families have been shown to be capable of inducing mesoderm¹. We sought to determine whether mesoderm induction by Laloo is mediated through the signal transduction pathways downstream of these ligands. The intracellular Smad proteins transduce signals from activated TGF- β receptors; Smad4 has a central role in this process⁹. A truncated Smad4 molecule (tSmad4) has been shown to block mesoderm induction by both Smad1 and Smad2, signal transducers of the bone morphogenetic proteins and activin, respectively⁹. To test whether mesoderm induction by Laloo requires signalling through the Smad pathway, we co-expressed Laloo and tSmad4. As expected, co-injection of tSmad4 inhibits mesoderm induction by Smad2 (Fig. 3e, lanes 4 and 5). In contrast, tSmad4 does not block induction by Laloo (Fig. 3e, lanes 1 and 2). Thus, we conclude that the induction of mesoderm by Laloo acts downstream, or independently, of the Smad proteins.

The other intracellular pathway known to be involved in mesoderm induction acts through stimulation of the Ras/MAP kinase cascade, downstream of the FGF receptor^{10–17}. To determine whether Laloo induces mesoderm independently of the FGF pathway, we first co-expressed Laloo and a dominant-inhibitory form of Ras¹¹. Dominant-inhibitory Ras entirely blocks mesoderm induction by FGF (Fig. 3e, lanes 13 and 14), and also blocks induction by Laloo (Fig. 3e, lanes 9 and 11). This result indicates that mesoderm induction by Laloo requires signalling through the wild-type Ras protein. We then challenged Laloo activity with a truncated form of the FGF receptor (XFD), also shown to act as a dominant-inhibitory molecule¹⁰. We reasoned that, as a putative intracellular signalling molecule, Laloo might bypass an inhibition by XFD at the cell surface. XFD blocks *Xbra* and *Xwnt8* induction by FGF, as expected (Fig. 3e, lanes 21 and 22). Interestingly, XFD also blocks the induction of *Xbra* and *Xwnt8* by Laloo (Fig. 3e, lanes 17 and 19). These results indicate that Laloo is either part of the FGF pathway or in a parallel pathway that requires signalling through the FGF receptor and Ras.

All Src-family proteins contain a C-terminal tyrosine that, when phosphorylated, markedly inhibits the activity of the protein². To examine whether similar regulation of Laloo might occur during early *Xenopus* development, we constructed a mutant form of Laloo in which we replaced the putative negative regulatory tyrosine residue with phenylalanine (Y492F). Y492F is indeed a more potent mesoderm inducer than wild-type Laloo. At midgastrula stages, 50 pg of Y492F RNA is sufficient to induce *Xbra* and *Xwnt8* (Fig. 4a, lanes 1–4), whereas 250 pg of *laloo* RNA is required to induce expression of these markers (Fig. 3a, lanes 1–4). At late neurula stages, 50 pg of Y492F RNA induces *HoxB9* expression, and 250 pg induces expression of *muscle actin* (Fig. 4b, lanes 1–4); 250 pg and 2 ng, respectively, of *laloo* RNA is required to induce *HoxB9* and *muscle actin* expression (Fig. 3b, lanes 1–4). Thus, the mesoderm-inducing activity of Laloo is modulated through a C-terminal tyrosine residue. As shown earlier, both dominant inhibitory Ras (Fig. 4c, lanes 1 and 3) and XFD (Fig. 4c, lanes 1 and 4) block mesoderm induction by wild-type Laloo. On the other hand, while mesoderm induction by Y492F is blocked by dominant-inhibitory Ras (Fig. 4c, lanes 2 and 5), it is largely unaffected by

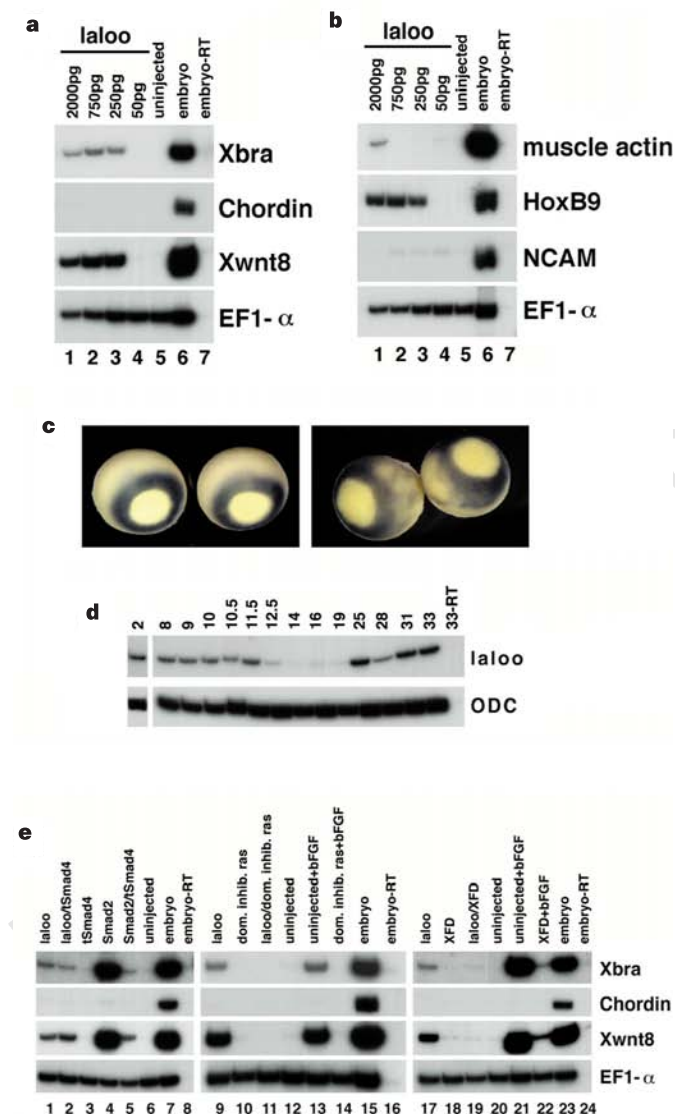


Figure 3 Ectopic Laloo induces mesoderm. RT-PCR analysis of animal caps dissected at late blastula stages and cultured until the stages listed. EF1- α is used as a loading control²⁹. The –RT lane contains all reagents except reverse transcriptase, and is used as a negative control. **a**, Analysis of animal caps cultured until midgastrula stages. Doses of *laloo* RNA over 2 ng were lethal. **b**, Analysis of animal caps cultured until late neurula stages. **c**, Whole-mount *in situ* hybridization of midgastrula stage embryos using an antisense *Xbra* probe. The embryos in the right panel were injected with 800 pg *laloo* RNA in the animal pole. **d**, Timing of *laloo* expression during early development. Ornithine decarboxylase (ODC) is used as a loading control³⁰. **e**, Induction of mesoderm by Laloo is unaffected by inhibition of Smad1 and Smad2, but is blocked by inhibition of the FGF signalling pathway. Animal caps were cultured until midgastrula stages. 750 pg each of *laloo*, *Smad2* and *tSmad4* RNA, 1.0 ng dominant-inhibitory Ras RNA and 1.5 ng XFD RNA were injected, as listed. Basic FGF (bFGF) was added to a final concentration of 25 ng ml^{–1}.

the co-expression of XFD (Fig. 4c, lanes 2 and 6). These results indicate that the requirement for FGF receptor activity is mediated through Tyr 492 of Laloo; furthermore, these data implicate Laloo as an integral component of the FGF signal transduction pathway, acting downstream of the FGF receptor and upstream of Ras.

In cell culture studies, kinase-defective mutants of Src-family members have been shown to act as dominant-negative molecules, blocking signalling through the wild-type kinase². We constructed a kinase-defective Laloo mutant by disrupting the putative ATP phosphotransferase site². This mutant, K259E, does not induce either *Xbra* or *Xwnt8* in mid-gastrula ectoderm explants (Fig. 4d, lane 2), demonstrating that mesoderm induction by Laloo requires a functional kinase domain. K259E expression inhibits mesoderm induction by Laloo at four-fold concentrations over wild-type (Fig. 4d, lanes 1 and 3); moreover, this mutant can inhibit mesoderm induction by both bFGF and activin in animal cap assays. Expression of K259E blocks induction of both *Xbra* and *Xwnt8* by bFGF (Fig. 4d, lanes 6 and 8), and blocks induction of *Xbra*, but not *Xwnt8* or *chordin*, by activin (Fig. 4d, lanes 10 and 12). Coexpression of Laloo moderately but consistently rescues mesoderm induction by FGF (Fig. 4d, lanes 6 and 7) and activin (Fig. 4d, lanes 10 and 11). This suggests that K259E acts by competition with wild-type Laloo. Rescue can also be achieved by coexpression of constitutively active Ras (not shown; ref. 11), supporting the placement of Laloo upstream of Ras in mesoderm induction. Marginal zone injections of K259E result in a reduction of trunk and tail structures *in vivo*

similar to that seen after injection of other molecules that disrupt the FGF signal transduction pathway^{10,12,14–16,18}; coexpression of either Laloo or constitutively active Ras partly rescues axis formation (Fig. 4e). These results indicate that Laloo, or a related factor inhibited by K259E, is required for mesoderm induction by FGF or activin and for normal development of the body axis.

Our results are consistent with a model placing Laloo as an essential intermediate in the FGF signalling pathway, downstream of the FGF receptor and upstream of Ras. The block of activin-induced *Xbra* expression by K259E suggests that Laloo might also mediate aspects of activin signalling; we believe, however, that this inhibition is indirect. Although FGF signalling is required for the full range of induction by activin, some activin-inducible genes are 'FGF-independent'; for example, *Xwnt8* is induced in XFD-expressing animal caps treated with activin^{13,19}. We have demonstrated both that K259E fails to block the induction of *Xwnt8* by activin and that Laloo activity is unaffected by coexpression of tSmad4; thus, the inhibition of activin-induced *Xbra* expression by K259E is likely to be secondary to a block of FGF signalling.

Experiments with XFD and the hyperactive Laloo construct Y492F suggest that mesoderm induction by Laloo requires the FGF-mediated dephosphorylation of Y492. Interestingly, inhibition of SH2-containing protein tyrosine phosphatase 2 (SH-PTP2) blocks mesoderm induction by FGF²⁰. Thus, *in vivo*, the activated FGF receptor might activate SH-PTP2, or a related phosphatase, which in turn could activate Laloo via dephosphorylation of Y492.

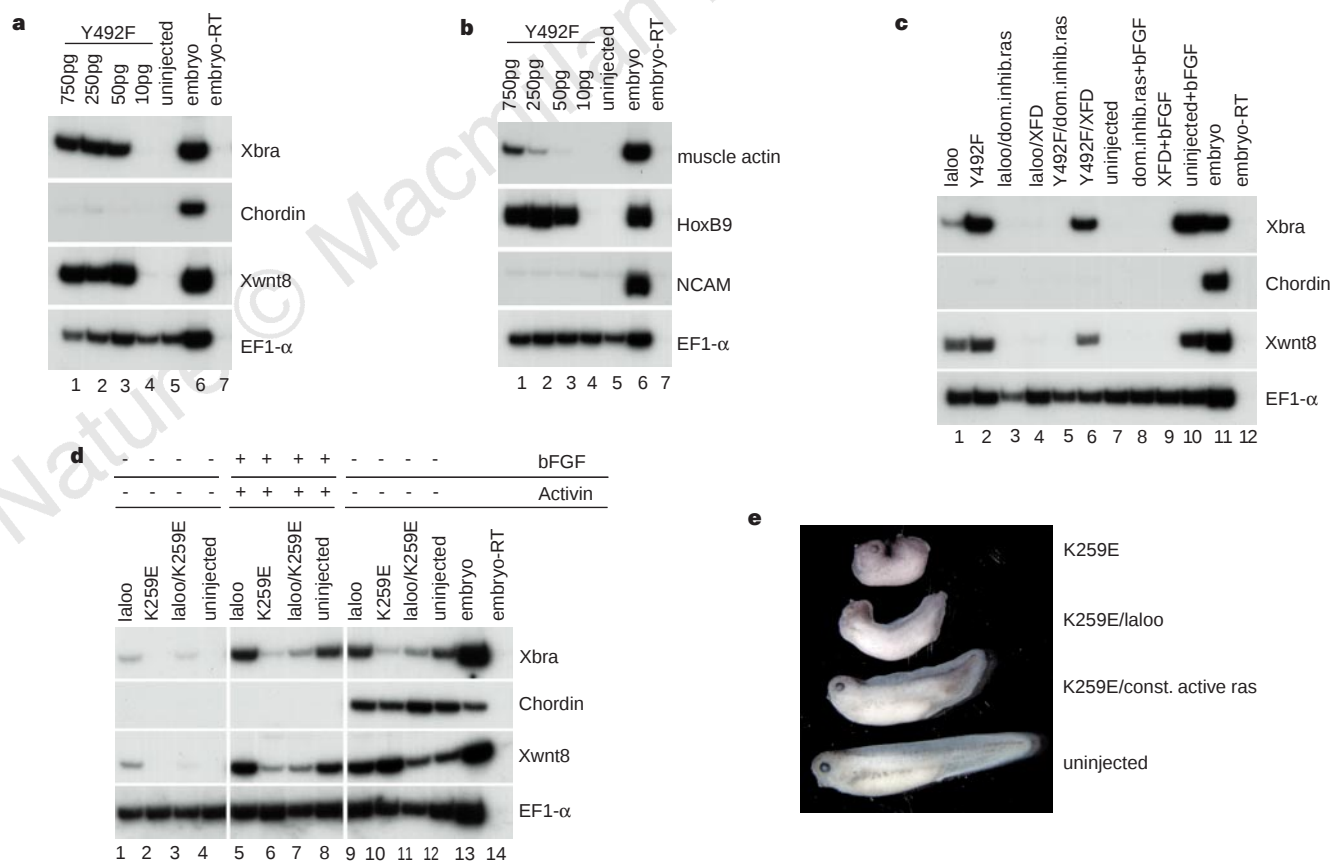


Figure 4 Y492F bypasses inhibition by XFD, and K259E inhibits the activity of mesoderm-inducing growth factors and normal development. Animal caps were cultured until midgastrula (**a, c, d**) or late neurula (**b**) stages. **a, b**, Y492F is a more potent mesodermal inducer than wild-type *laloo*. **c**, A hyperactive *laloo* mutant bypasses inhibition by the truncated FGF receptor. 750 pg *laloo*, 250 pg Y492F, 1.0 ng dominant-inhibitory Ras and 1.5 ng XFD were injected. **d**, K259E inhibits mesoderm induction by bFGF and activin. 800 pg *laloo* and 3.2 ng K259E RNA were injected. Inhibition by K259E required doses of 3.2 ng or more. **e**, K259E perturbs normal development of the body axis. Lateral view of stage 37 embryos;

2 ng K259E, 400 pg *laloo* and 40 pg activated *Ras* RNA were injected radially, as listed, at early cleavage stages. Top, embryo injected with K259E RNA. 81% of these embryos display a failure of blastopore closure and a complete loss of tail structures ($n = 78$). Second from top, embryo injected with both K259E and *laloo* RNA (56% with posterior truncations; $n = 70$). These embryos often show a reduction of anterior structures. Second from bottom, embryo injected with both K259E and constitutively active *Ras* RNA (33% with posterior truncations, $n = 64$). Bottom, un.injected control embryo.

Our results also suggest that the activity of ectopic *Laloo* is dependent upon a basal level of signalling through the FGF receptor. In support of this, it has been shown that FGF receptor-dependent MAP kinase activity is present throughout the early *Xenopus* embryo²¹. It has also been demonstrated that an autoregulatory loop involving FGF is required for mesoderm maintenance^{18,22,23}; Y492F might thus generate ectopic mesoderm in the presence of XFD by bypassing the requirement for continued signalling through the FGF receptor.

The positioning of Src-related factors between the FGF receptor and Ras has precedents in other studies. Src-family kinases have previously been shown to transmit signals through Ras; the molecular interactions proposed to link these factors include phosphorylation of Shc and/or Ras GTPase-activating protein². Functional association between Src-related proteins and receptor tyrosine kinases has also been demonstrated²⁴. Our studies do not, however, address whether signalling upstream of *Laloo* is mediated solely through the FGF receptor; *Laloo* might also receive input from other cell-surface receptors. Furthermore, redundancy has been demonstrated among Src-family genes in other biological contexts²; mesoderm induction *in vivo* might thus involve additional Src-related factors. Although a constitutively active form of chicken Src has been reported not to induce mesoderm in *Xenopus* explants¹³, this apparent difference might be due to cross-species incompatibility as well as to differences in assay conditions. Efforts are currently in progress to address these possibilities. This report provides strong evidence of a critical role for a Src-family kinase in early vertebrate development, acting as a required component of the FGF signal transduction pathway. □

Methods

Expression library construction and screening. Early gastrula (stage 10) *Xenopus laevis* embryos were homogenized with RNazol B solution and processed in accordance with the manufacturer's instructions (Tel-Test). 4.5 µg poly(A)⁺ RNA was selected from 2.4 mg total RNA using the Oligotex mRNA midi kit (Qiagen). cDNA synthesis and linker addition was performed using the Superscript II unidirectional kit (Gibco BRL). After second-strand synthesis, cDNAs were size-selected by gel filtration (average size 2.0 kb) and directionally subcloned into the *SalI* and *NotI* sites of a modified pCS2 vector. 2 × 10⁶ transformants were obtained after electroporation in ElectroMAX DH10B cells (Gibco BRL).

The library was plated to obtain initial fractions of approximately 200 clones. Similar functional screens in *Xenopus*, using small pools of RNAs, have been described elsewhere²⁵. For subsequent sib selection, 10 pools of fivefold fewer clones were screened (e.g. 10 × 40, 10 × 8). Pooled plasmid DNA was isolated using the QIAprep spin miniprep kit (Qiagen) and linearized with *AscI*. Capped RNA was synthesized using the mMessage mMachine kit (Ambion). Embryos were injected with 10 nl/blastomere of 0.5 mg ml⁻¹ RNA into animal poles of both blastomeres at the two-cell stage, and cultured until tadpole stages.

RNA preparation, explant dissection and cell culture. *laloo* RNA and all mutant derivatives contained the *laloo* open reading frame, as well as 84 nucleotides of 5' and 34 nucleotides of 3' untranslated sequence. All mRNA was synthesized *in vitro* in the presence of cap analogue using the mMessage mMachine kit. Microinjection, explant dissection and culture were performed as described²⁶. Recombinant bFGF was obtained from Boehringer Mannheim.

Whole-mount immunohistochemistry, β-Gal detection and *in situ* hybridization. Whole-mount antibody staining was performed as described²⁶. The 6F11 antibody was used at 1:1 dilution, and the 12/101 antibody was used at 1:500 dilution. Secondary antibody was a donkey anti-mouse IgG coupled to horseradish peroxidase (Jackson Laboratories) and was used at 1:250 dilution. Colour reactions were performed using the DAB and Vector SG kits (Vector Laboratories). Whole-mount β-Gal detection and hybridization *in situ* were performed as described⁴. The antisense *Xbra* probe was synthesized in the presence of digoxigenin-11-UTP (Boehringer Mannheim).

RT-PCR. RT-PCR was performed as described⁹. Primers constructed for this study were as follows: *laloo* (forward, 5'-TGGCTCTGTACTGTGATC; reverse, 5'-GTCATACAAAGCCAGCAG); *chordin* (forward, 5'-CAGTCAGATGGAG-CAGGATC; reverse, 5'-AGTCCCATTGCCCGAGTTGC); *ODC* (forward, 5'-

AATGGATTTTCAGAGACCA; reverse, 5'-CCAAGGCTAAAGTTGCG). All other primer sequences were as described²⁶. PCR for *laloo*, *Xwnt8*, *HoxB9* and *NCAM* were performed for 25 cycles; PCR for *EF1-α*, *Xbra*, *chordin*, *muscle actin* and *ODC* were performed for 21 cycles.

Preparation of *laloo* mutant constructs. The *laloo* mutants K259E and Y492F were generated by PCR. For K259E, we introduced a point mutation (A → G) that resulted in a lysine (AAA) to glutamic acid (GAA) mutation in the resulting construct. The oligonucleotide used for this mutagenesis was as follows: 5'-GTA GAA ACA ATG AAG CCA GGC AGC. For Y492F, we introduced a point mutation (A → T) that resulted in a tyrosine (TAC) to phenylalanine (TTC) mutation in the resulting construct. The complementary strand oligonucleotide thus includes a T → A mutation: 5'-TTA AGG TTG TGC CTG GAA CTG.

Received 10 March; accepted 26 June 1998.

- Klein, P. S. & Melton, D. A. Hormonal regulation of embryogenesis: the formation of mesoderm in *Xenopus laevis*. *Endocr. Rev.* **15**, 326–341 (1994).
- Brown, M. T. & Cooper, J. A. Regulation, substrates and functions of src. *Biochim. Biophys. Acta* **1287**, 121–149 (1996).
- Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D. & Herrmann, B. G. Expression of a *Xenopus* homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79–87 (1991).
- Smith, W. C. & Harland, R. M. Injected *Xwnt-8* RNA acts early in *Xenopus* embryos to promote formation of a vegetal dorsalizing center. *Cell* **67**, 753–765 (1991).
- Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L. K. & DeRobertis, E. M. *Xenopus* chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* **79**, 779–790 (1994).
- Wright, C. V. E., Morita, E. A., Wilkin, D. J. & DeRobertis, E. M. The *Xenopus* *X1Hbox6* homeo protein, a marker of posterior neural induction, is expressed in proliferating neurons. *Development* **109**, 225–234 (1990).
- Kintner, C. R. & Melton, D. A. Expression of *Xenopus* NCAM RNA in ectoderm is an early response to neural induction. *Development* **99**, 311–325 (1987).
- Mohun, T. J., Brennan, S., Dathan, N., Fairman, S. & Gurdon, J. B. Cell type-specific activation of actin genes in the early amphibian embryo. *Nature* **311**, 716–721 (1984).
- Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massague, J. Partnership between DPC4 and SMAD proteins in TGF-β signalling pathways. *Nature* **383**, 832–836 (1996).
- Amaya, E., Musci, T. J. & Kirschner, M. W. Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in *Xenopus* embryos. *Cell* **66**, 257–270 (1991).
- Whitman, M. & Melton, D. A. Involvement of p21^{ras} in *Xenopus* mesoderm induction. *Nature* **357**, 252–255 (1992).
- MacNicol, A. M., Muslin, A. J. & Williams, L. T. Raf-1 kinase is essential for early *Xenopus* development and mediates the induction of mesoderm by FGF. *Cell* **73**, 571–584 (1993).
- LaBonne, C. & Whitman, M. Mesoderm induction by activin requires FGF-mediated intracellular signals. *Development* **120**, 463–472 (1994).
- LaBonne, C., Burke, B. & Whitman, M. Role of MAP kinase in mesoderm induction and axial patterning during *Xenopus* development. *Development* **121**, 1475–1486 (1995).
- Gotoh, Y., Masuyama, N., Suzuki, A., Ueno, N. & Nishida, E. Involvement of the MAP kinase cascade in *Xenopus* mesoderm induction. *EMBO J.* **14**, 2491–2498 (1995).
- Umbhauer, M., Marshall, C. J., Mason, C. S., Old, R. W. & Smith, J. C. Mesoderm induction in *Xenopus* caused by activation of MAP kinase. *Nature* **376**, 58–62 (1995).
- Northrop, J. et al. BMP-4 regulates the dorsal-ventral differences in FGF/MAPKK-mediated mesoderm induction in *Xenopus*. *Dev. Biol.* **172**, 242–252 (1995).
- Amaya, E. & Kroll, K. L. Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development* **122**, 3173–3183 (1996).
- Cornell, R. A. & Kimmel, D. Activin-mediated mesoderm induction requires FGF. *Development* **120**, 453–462 (1994).
- Tang, T. L., Freeman, R. M. Jr, O'Reilly, A. M., Neel, B. G. & Sokol, S. Y. The SH2-containing protein-tyrosine phosphatase SH-PTP2 is required upstream of MAP kinase for early *Xenopus* development. *Cell* **80**, 473–483 (1995).
- LaBonne, C. & Whitman, M. Localization of MAP kinase activity in early *Xenopus* embryos: implications for endogenous FGF signaling. *Dev. Biol.* **183**, 9–20 (1997).
- Isaacs, H. V., Pownall, M. E. & Slack, J. M. W. eFGF regulates *Xbra* expression during *Xenopus* gastrulation. *EMBO J.* **13**, 4469–4481 (1994).
- Schulte-Merker, S. & Smith, J. C. Mesoderm formation in response to Brachyury requires FGF signaling. *Curr. Biol.* **5**, 62–67 (1995).
- Parsons, J. T. & Parsons, S. J. Src family protein tyrosine kinases: cooperating with growth factor and adhesion signaling pathways. *Curr. Opin. Cell Biol.* **9**, 187–192 (1997).
- Lustig, K. D., Kroll, K. L., Sun, E. E. & Kirschner, M. W. Expression cloning of a *Xenopus* T-related gene (*Xombi*) involved in mesodermal patterning and blastopore lip formation. *Development* **122**, 4001–4012 (1996).
- Hemmati-Brivanlou, A. & Melton, D. A. Inhibition of activin receptor signaling promotes neuralization in *Xenopus*. *Cell* **77**, 273–281 (1994).
- Harris, W. A. & Hartenstein, V. Neuronal determination without cell division in *Xenopus* embryos. *Neuron* **6**, 499–515 (1991).
- Kintner, C. R. & Brockes, J. P. Monoclonal antibodies identify blastemal cells derived from dedifferentiating muscle in newt limb regeneration. *Nature* **308**, 67–69 (1984).
- Krieg, P., Varnum, S., Wormington, M. & Melton, D. A. The mRNA encoding elongation factor 1α (EF-1α) is a major transcript at the mid-blastula transition in *Xenopus*. *Dev. Biol.* **133**, 93–100 (1989).
- Bassez, T., Paris, J., Omilli, E., Dorel, C. & Osborne, H. B. Post-transcriptional regulation of ornithine decarboxylase in *Xenopus laevis* oocytes. *Development* **110**, 955–962 (1990).

Acknowledgements. We thank M. Whitman and E. Amaya for providing us with the Ras and XFD constructs, respectively; W. Harris for the 6F11 antibody; and members of the laboratory, P. Wilson and H. Hanafusa, for critical reading of the manuscript. E.C. received partial financial support from Programma Scambi Internazionali and Piano Bilaterale of CNR of Rome. A.H.-B. is a Merck and a McKnight scholar. This work was supported by NIH grant HD 32105-01 and by grants from the Klingenstein and Merck Foundations to A.H.-B.

Correspondence and requests for materials should be addressed to A.H.-B. (e-mail: brvnlou@rockvax.rockefeller.edu). The GenBank accession number for *laloo* is AF081803.

Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF- β -induced transcription

Ying Zhang, Xin-Hua Feng & Rik Derynck

Departments of Growth and Development, and Anatomy, Programs in Cell Biology and Developmental Biology, University of California at San Francisco, San Francisco, California 94143-0640, USA

Smad proteins transduce signals for transforming growth factor- β (TGF- β)-related factors¹. Smad proteins activated by receptors for TGF- β form complexes with Smad4. These complexes are translocated into the nucleus and regulate ligand-induced gene transcription²⁻⁴. 12-O-tetradecanoyl-13-acetate (TPA)-responsive gene promoter elements (TREs) are involved in the transcriptional responses of several genes to TGF- β (refs 5-8). AP-1 transcription factors, composed of c-Jun and c-Fos, bind to and direct transcription from TREs, which are therefore known as AP-1-binding sites⁹. Here we show that Smad3 interacts directly with the TRE and that Smad3 and Smad4 can activate TGF- β -inducible transcription from the TRE in the absence of c-Jun and c-Fos. Smad3 and Smad4 also act together with c-Jun and c-Fos to activate transcription in response to TGF- β ,

through a TGF- β -inducible association of c-Jun with Smad3 and an interaction of Smad3 and c-Fos. These interactions complement interactions between c-Jun and c-Fos, and between Smad3 and Smad4. This mechanism of transcriptional activation by TGF- β , through functional and physical interactions between Smad3-Smad4 and c-Jun-c-Fos, shows that Smad signalling and MAPK/JNK signalling converge at AP-1-binding promoter sites.

Several AP-1-regulated promoters are transcriptionally induced by TGF- β (refs 5-8, 10) and Smad2 or Smad3 with Smad4 (refs 2, 3, 11). We therefore tested the effect of TGF- β and Smad proteins on transcription from a synthetic reporter, TRE-Luc¹², which contains four tandem AP-1-binding sites from the collagenase I promoter¹³. Transcription from this reporter in TGF- β -responsive Mv1Lu cells was induced by TGF- β , occurred as early as 2 h after TGF- β treatment (data not shown), and reached 15-20-fold induction after 12 h (Fig. 1a). Among the Smad proteins tested, only Smad3 induced moderate ligand-independent transcription and enhanced the response to TGF- β (Fig. 1a). Smad3 and Smad4 acted together to induce marked ligand-independent and ligand-dependent transcriptional activation, consistent with the required cooperation of both Smad proteins in a heteromeric complex^{3,14}. Co-expression of Smad2 and Smad4 also conferred transcriptional activation, albeit to a lesser extent than Smad3 and Smad4, whereas the promoter was not responsive to Smad1 and Smad4.

We next studied the roles of c-Jun or c-Fos in the transcriptional activity of Smad3 and Smad4. In F9 cells, which lack c-Jun and c-Fos^{15,16} yet have endogenous Smad3 and Smad4 (data not shown),

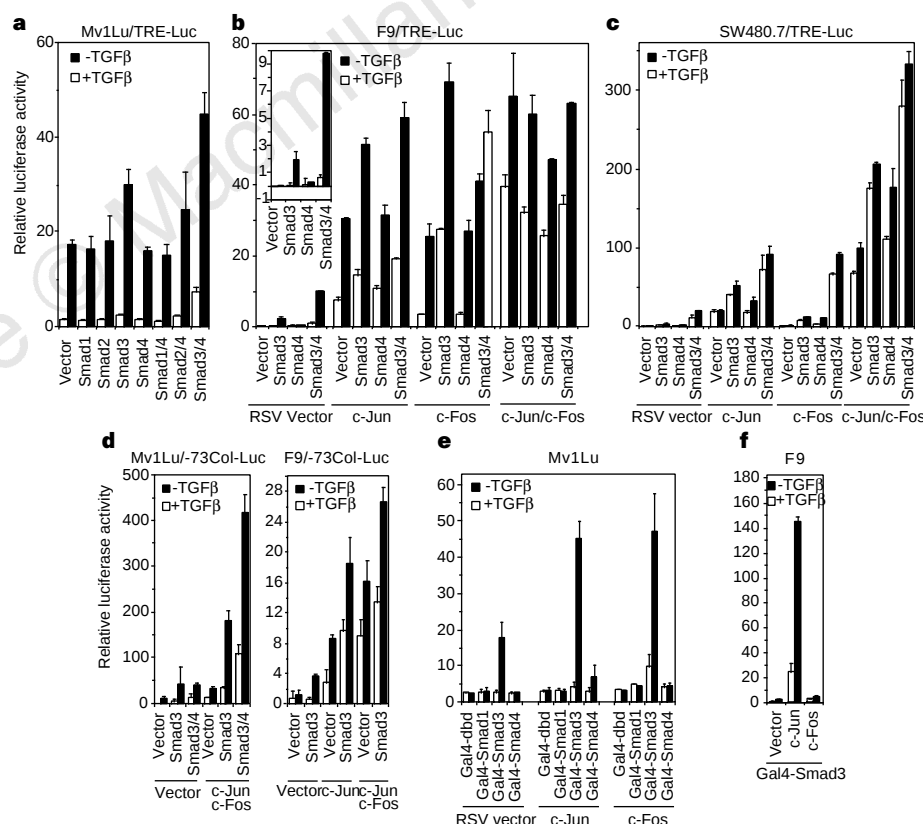


Figure 1 Smad3/Smad4 and c-Jun-c-Fos cooperate to induce transcription from the AP-1 promoter. We measured the effects of Smads, c-Jun, and c-Fos on luciferase expression from AP-1 promoters in the presence or absence of TGF- β . Values are relative to control cells in the absence of TGF- β . The reporter plasmid TRE-Luc was used in **a-c**, whereas the -73Col-luc reporter was used in **d-f**. **a**, TGF- β -responsive Mv1Lu cells. **b**, F9 cells. The inset shows on an extended scale the effects of Smad3 and/or Smad4 on transcription. **c**, TGF- β -unresponsive, Smad4-defective SW480.7 cells. Note the difference in scale compared with **b**. **d**, Smad3 and Smad3-Smad4 synergized with c-Jun and c-Fos to induce transcription from

the -73Col-luc reporter, which contains a single AP-1-binding site, in Mv1Lu and F9 cells. **e**, **f**, we used pFR-Luc, with its five Gal4 DNA-binding sequences, as a reporter for transcription mediated by Gal4-fusion proteins. Values are relative to control cells transfected with Gal4-dbd (DNA-binding domain) in the absence of TGF- β . **e**, In Mv1Lu cells, the transcription by Gal4-Smad3, but not Gal4-Smad4, was induced by TGF- β and further enhanced in a TGF- β -dependent way by c-Jun or c-Fos. **f**, In F9 cells, c-Jun increased transcription by Gal4-Smad3, which was strongly enhanced by TGF- β . In contrast with Mv1Lu cells, co-expression of c-Fos alone had little effect on transcription by Gal4-Smad3 in F9 cells.

Smad3 overexpression activated transcription in response to TGF- β , and Smad4 synergistically enhanced this activity in the absence of c-Jun and c-Fos (Fig. 1b, inset). c-Jun and c-Fos conferred a high level of TGF- β -inducible transcription from this promoter (Fig. 1b). Smad3, but not Smad4, acted with either c-Jun or c-Fos to induce strong ligand-independent transcription, which was further enhanced by TGF- β . Finally, the high level of TGF- β -independent transcription mediated by c-Jun–c-Fos was further enhanced by TGF- β . The strong transcriptional activation by co-expressed c-Jun and c-Fos was not further increased by Smad3 and/or Smad4, possibly because a saturating level of stimulation was already reached (Fig. 1b). We also observed a synergy of Smad3/4 with c-Jun or c-Fos in TGF- β -responsive Mv1Lu cells (data not shown).

In SW480.7 cells, which lack Smad4 and, consequently, responsiveness to TGF- β (ref. 3), and in contrast with F9 and Mv1Lu cells, c-Fos alone did not allow transcription from the TRE (Fig. 1c), which is consistent with the inability of c-Fos to bind DNA directly⁹ and indicates that the effect of c-Fos in F9 and Mv1Lu cells depended on endogenous Smad3 and Smad4. The synergy of Smad3 and Smad4 with c-Jun and c-Fos was more apparent in SW480.7 cells than in F9 and Mv1Lu cells, and co-expression of Smad4 enhanced the already high level of synergy of Smad3 with c-Jun or c-Fos or both. Finally, Smad3–Smad4 acted together with c-Jun–c-Fos at the 73-base-pair (bp) human collagenase I promoter with its single consensus TRE¹³ (Fig. 1d), consistent with the results using TRE-luc (Fig. 1a–c). These results indicate that the transcriptional cooperativity between Smad3–Smad4 and c-Jun–c-Fos also occurs at a single AP1-binding site.

We also characterized the synergy between Smad3–Smad4 and

c-Jun–c-Fos in Gal4 transactivation assays, in which we assessed the effects of c-Jun or c-Fos on the ability of Smad proteins, fused to a Gal4 DNA-binding domain, to drive transcription from a promoter with five tandem Gal4-binding sites (Fig. 1e, f). In Mv1Lu cells, TGF- β activated transcription by Gal4–Smad3, but not by Gal4–Smad1, and c-Jun and c-Fos enhanced the TGF- β -dependent transcription by Smad3 (Fig. 1e). In F9 cells, c-Jun, but not c-Fos, enhanced transcription by Gal4–Smad3 (Fig. 1f). This indicates that the effect of c-Fos on Gal4–Smad3 activity in Mv1Lu cells (Fig. 1e) is due to its synergy with endogenous c-Jun, whereas c-Jun does not require c-Fos to transactivate Smad3. In contrast to Smad3, the activity of Gal4–Smad4 was not induced by TGF- β and was increased only minimally by c-Jun or c-Fos (Fig. 1e). As Smad4 alone does not activate transcription^{14,17}, this minimal increase is probably due to synergy with endogenous Smad3.

The transcriptional cooperation between Smad3 and c-Jun correlates with a physical interaction between these proteins. In Mv1Lu cells, mammalian two-hybrid assays showed only low-level interactions of Smad3 with c-Jun or c-Fos in the absence of TGF- β . Addition of TGF- β induced a strong interaction of Smad3 with c-Jun and a weaker interaction of Smad3 with c-Fos (Fig. 2a). In F9 cells, the low-level interaction of Smad3 with c-Fos was not increased by TGF- β , whereas, as in Mv1Lu cells, the interaction of Smad3 and c-Jun was strongly enhanced by TGF- β (Fig. 2b). These results indicate that the TGF- β -induced interaction of Smad3 with c-Fos in Mv1Lu cells may be due to stabilization by endogenous c-Jun. In transfected cells expressing Smad3, c-Jun co-immunoprecipitated with Smad3 following TGF- β -receptor activation (Fig. 2c, d). Only a small amount of c-Fos co-precipitated with Smad3, and

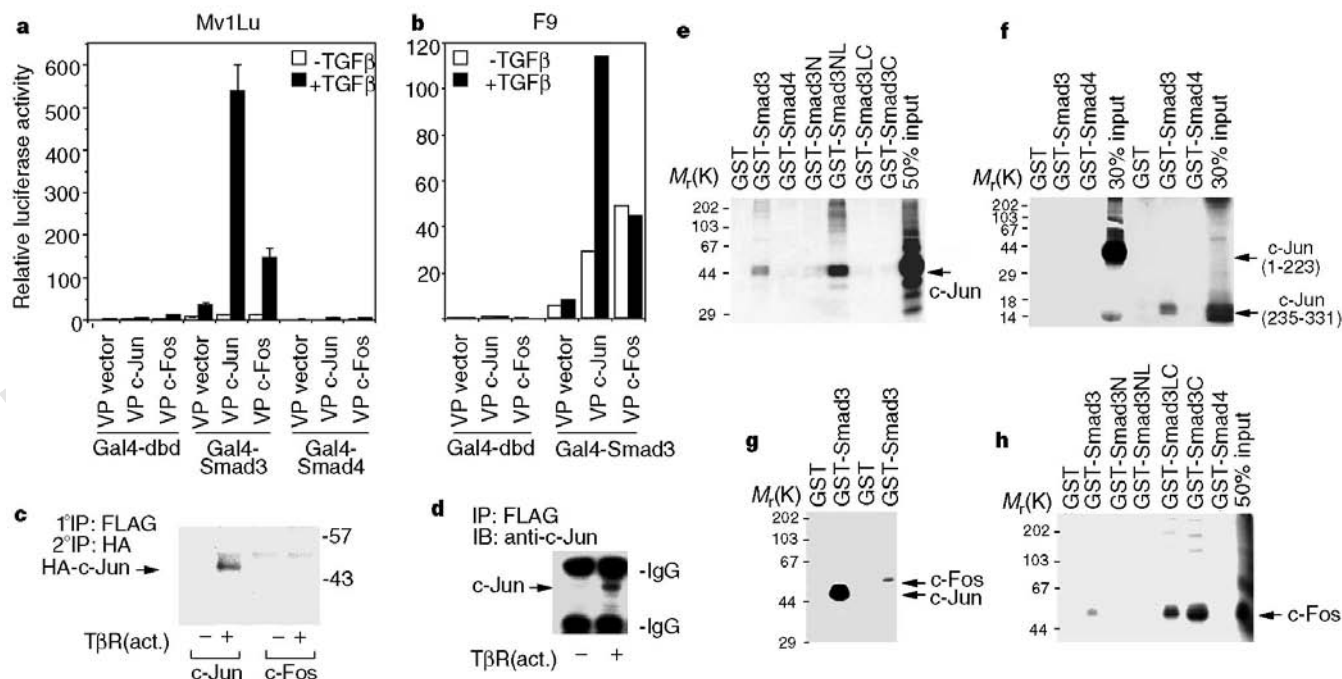


Figure 2 Interaction between Smad3 and c-Jun or c-Fos. **a**, Mammalian two-hybrid assays in Mv1Lu cells reveal TGF- β -inducible interactions between c-Jun or c-Fos and Smad3, but not Smad4. Smad3 and Smad4 were expressed as fusions with the GAL4 DNA-binding domain and c-Jun and c-Fos were expressed as fusions with the VP16 transactivation domain. Interactions were scored as transcription from pFR-luc; dbd represents the DNA-binding domain as control. **b**, Mammalian two-hybrid assays were performed as in **a**, except that F9 cells were used. Note the difference in scale compared with **a**. **c**, Ligand-dependent association of transfected c-Jun and Smad3. Flag-tagged Smad3 and HA-tagged c-Jun or c-Fos were co-expressed in ³⁵S-labelled Cos-1 cells and sequential immunoprecipitations (IP), first with anti-Flag and, following dissociation, with anti-HA antiserum, were performed. c-Jun precipitated with Smad3 only following

receptor (T β R) activation. The upper band shows co-precipitated Smad3, which, in the case of the c-Fos immunoprecipitations, may comigrate with c-Fos. **d**, Endogenous c-Jun interacts with Smad3 following receptor (T β R) activation. Flag-tagged Smad3, expressed in Cos-1 cells, was immunoprecipitated and the association of c-Jun with Smad3 was visualized by western blotting using anti-c-Jun antibodies. **e**, Smad3 interacts directly with c-Jun. The c-Jun-interaction domain of Smad3 is localized at its N-L segment. **f**, c-Jun interacts with Smad3 through its C-terminal bZIP domain. **g**, The interaction of c-Fos with Smad3 is much weaker than the interaction of c-Jun with Smad3. Equal mounts of *in vitro*-translated ³⁵S-labelled c-Jun or c-Fos, with similar specific radioactivity, were incubated with GST or GST-Smad3. **h**, Direct association of c-Fos with the conserved C-domain of Smad3.

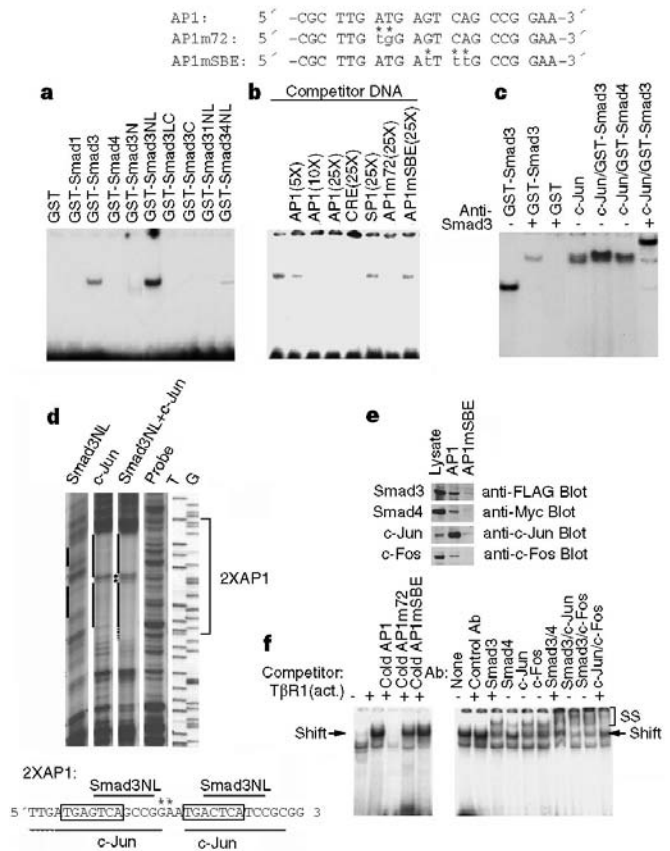


Figure 3 Smad3-Smad4 and c-Jun-c-Fos participate in a nucleoprotein complex with the consensus AP1-binding sequence. **a**, Direct binding of Smad3, but not Smad1 or 4, to the AP-1 consensus sequence shown at the top. We used semipurified GST or GST-fused Smad proteins in gel shift assays. The DNA-binding sequence of Smad3 is found in its N-L segment. The N-domain showed minimal DNA binding. The C-domain with or without the L segment did not bind DNA. **b**, Excess unlabelled AP-1 or CRE (agagattgccTGACGTCAgagagctag) probes, but not Sp1 (attcgatCGGGGCGGGCcgagc) probe, competes with GST-Smad3 for binding to the AP-1 sequence. A mutated AP-1 probe, AP1m72, which has mutations in the AP-1 consensus sequence and does not bind c-Jun, but not another mutant AP-1 probe, AP1mSBE, which has mutations in both AP1- and Smad3-binding sequences, competes with Smad3 for binding to the AP-1 sequence. **c**, GST-Smad3 without or with purified c-Jun was incubated with the [³²P]-AP-1 probe, and gel shift analyses were performed. Anti-Smad3 antibodies supershifted GST-Smad3-DNA and c-Jun-GST-Smad3-DNA complexes, indicating that c-Jun and Smad3 bind to the AP-1 probe together. **d**, Footprinting analysis of the interaction of Smad3 and/or c-Jun with the AP1-binding site. The protected sequences (determined from parallel sequencing reactions using ddT and ddG) are marked next to the lanes and on the sequence shown below. **e**, Smad3, Smad4, c-Jun and c-Fos interact with the AP1-binding consensus oligonucleotide. Immobilized AP-1 or mutated AP1mSBE oligonucleotides were incubated with cell lysates from Cos-1 cells transfected with expression plasmids for Flag-Smad3, Smad4-Myc, c-Jun, c-Fos and activated TGFβ receptors (TβRI (act.)). The affinity-purified complex was analysed by immunoblotting using the indicated antibodies. The lysate control lanes showed that all four proteins were expressed. **f**, Gel shift and supershift analyses using the [³²P]-AP-1 probe in nuclear extracts from 293 cells transfected with expression plasmids for Flag-Smad3, Smad4-Myc, HA-c-Jun or c-Fos, with or without activated TGFβ receptors. In the left panel, excess unlabelled AP-1 or mutant AP-1 probe was included in the reaction as indicated. In the right panel, antibodies were added as indicated. The TGF-β-dependent protein-DNA (shift) and supershifted (SS) complexes are indicated.

this was not always observed. These results show a ligand-dependent association of Smad3 with c-Jun, and indicate that receptor-induced phosphorylation and nuclear translocation of Smad3 may be essential for association of Smad3 with c-Jun.

To characterize the interaction of Smad3 with c-Jun and c-Fos, we used glutathione S-transferase (GST) binding assays. c-Jun associated directly with Smad3, but not with Smad4; c-Jun associated more efficiently with the N-L segment of Smad3 and not with the C-domain (Fig. 2e). Smad3 interacted with the carboxy-terminal segment of c-Jun, which contains the bZIP domain that mediates DNA binding and dimerization⁹, and not with its amino-terminal segment (Fig. 2f). Consistent with the weak interactions between c-Fos and Smad3 in mammalian cells (Fig. 2a, b), c-Fos interacted much less with Smad3 than did c-Jun; the c-Fos-Smad3 interaction was mediated by the C-domain of Smad3, and the L or N-L segment decreased this association (Fig. 2g, h). Together, our results indicate that decreased affinity of the C-domain for the N-domain¹⁸ following receptor-induced C-terminal phosphorylation of Smad3 exposes the N-L segment, allowing binding to c-Jun, and the C-domain, allowing binding to c-Fos. This complex can then activate transcription more efficiently.

As shown in gel mobility-shift assays, Smad3 associated directly with the AP-1 sequence from the collagenase I promoter, and this interaction was mediated by the N-L segment of Smad3 (Fig. 3a). The C- and L-C domains, which can activate transcription when fused to a DNA-binding domain^{14,19}, did not bind the AP-1 sequence. The N-L domain of Smad4 showed very weak binding to the AP-1 sequence and full-length Smad4 did not bind. Full-length Smad1 and its N-L domain did not bind the AP-1 sequence. The Smad3-DNA interaction competed with unlabelled AP-1 probe, but not with the unrelated Sp1-binding sequence GGCGGGG. The cyclic-AMP-response element (CRE) TGACGTCA, which differs from the TPA-responsive AP-1 site only in having one extra cytosine, also competed efficiently for Smad3 binding (Fig. 3b). TGF-β and Smad3 and Smad4 repress the promoter activity of the cyclin A promoter^{3,20}; this downregulation by TGF-β requires the ATF-1/CREB(CRE-binding protein)-binding CRE sequence²¹. Thus, direct binding of Smad3 to the CRE site may mediate this downregulation by TGF-β.

c-Jun homodimers and c-Jun-c-Fos dimers activate transcription through their ability to interact directly with the AP1-binding site⁹. Smad3 and c-Jun interacted individually with the AP-1 oligonucleotide, as expected. Incubation of both Smad3 and c-Jun with this oligonucleotide resulted in a more slowly migrating, more intense complex than the complex formed with c-Jun alone, and this complex was supershifted using an anti-Smad3 antibody. These results indicate that Smad3 and c-Jun can bind simultaneously to the TRE and cooperate to form a more stable complex than would be formed with either protein alone. Footprinting analysis showed that the N-L domain of Smad3 protected the sequence GTCAGCC, which overlaps with the AP1-binding sequence TGAGTCA¹³ (Fig. 3d) and resembles the Smad3/4-binding sequence GTCTAGNC^{22,23}. Incubation of Smad3 and c-Jun with the AP-1 oligonucleotide conferred a protection pattern that reproducibly differed to some extent from the c-Jun pattern and was distinct from the Smad3 pattern. Consistent with the different footprinting patterns, Smad3 and c-Jun showed differential sequence requirements for binding at the TRE. Thus, a mutant sequence, AP1m72, which fails to bind c-Jun¹³, still competed with the ³²P-labelled AP-1 probe for Smad3 binding, whereas another oligonucleotide, AP1mSBE, which also does not bind c-Jun¹³ because of mutations in both AP1- and Smad3-binding elements, did not compete with the AP-1 probe for Smad3 binding (Fig. 3b). Overlapping AP1 and Smad3-binding sequences, as found in several TGF-β-regulated promoters, may be important for TGF-β-induced transcription. Thus, not all AP1-binding sites may have equal affinities for Smad3, and this affinity

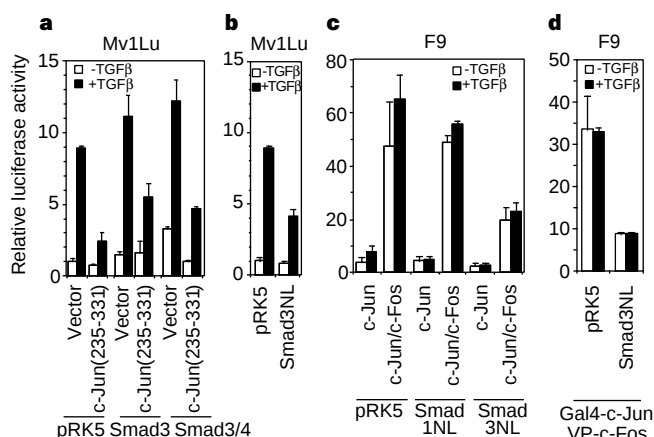


Figure 4 Dominant-negative inhibition of TGF- β - and Smad3/4-induced transcription from the AP1-binding site by truncated c-Jun or Smad3. **a**, Overexpression of the dominant-negative c-Jun mutant, c-Jun(235–331), inhibits transcriptional activation from the TRE promoter by TGF- β , Smad3 or Smad3–Smad4 in Mv1Lu cells. c-Jun(235–331) contains only the C-terminal bZIP domain of c-Jun. **b**, Overexpression of Smad3NL, that is, the N–L segment of Smad3, inhibits TGF- β -induced transcription from the TRE promoter in Mv1Lu cells. **c**, Overexpression of Smad3NL but not Smad1NL inhibits c-Jun- or c-Jun–c-Fos-induced transcription from the TRE promoter in F9 cells. **d**, Smad3NL inhibits the physical interaction between Gal4–c-Jun and VP–c-Fos as scored in mammalian two-hybrid assays.

may be determined by a few nucleotides that flank the AP1-binding heptanucleotide.

Smad3 and Smad4 also participated with c-Jun and c-Fos in a multimeric complex at the AP-1/Smad3-binding site *in vivo*. Oligonucleotides containing the AP-1/Smad3-binding sequence interacted with c-Jun, c-Fos, Smad3 and Smad4 in transfected cell lysates (Fig. 3e), although only Smad3 and c-Jun can bind this sequence directly. The interaction of c-Jun–c-Fos and Smad3–Smad4 at the AP-1/Smad-binding sequence was further supported by supershift analyses using different antibody combinations (Fig. 3f). TGF- β -receptor activation induced the formation of a DNA–protein complex, which was absent in unstimulated cells. This TGF- β -dependent complex competed with excess unlabelled AP-1/Smad3-binding oligonucleotide, but competed inefficiently with two different mutated oligonucleotides that do not bind c-Jun, suggesting a key role of c-Jun in the formation of this complex (Fig. 3f). The TGF- β -inducible complex was supershifted using antibodies against Smad3, Smad4, c-Jun or c-Fos (Fig. 3f). The incomplete supershift of the DNA–protein complex by the antibodies is not surprising, as several other bZIP transcription factors, including TGF- β -inducible JunB, can form complexes with different transcriptional activities at the AP1-binding site²⁴. A combination of two different antibodies resulted in a supershift band of even slower electrophoretic mobility (Fig. 3f). The interaction data and supershift analyses shown in Figs 2 and 3, combined with the functional interactions of c-Jun–c-Fos and Smad3–Smad4 (Fig. 1), indicate that all four components, that is, Smad3, Smad4, c-Jun and c-Fos, form a multimeric protein complex at the AP-1/Smad-binding sequence.

We studied the role of c-Jun in TGF- β signalling further by using a truncated version of c-Jun. c-Jun(235–331) contains the bZIP domain of c-Jun and acts as a dominant-negative inhibitor of AP-1 activity²⁵, presumably by sequestering c-Jun and c-Fos or by occupying the AP-1 DNA-binding site. c-Jun(235–331) inhibited TGF- β - and Smad3/4-induced transcription from TRE-luc (Fig. 4a). As this segment of c-Jun interacts with Smad3, c-Jun(235–331) may inhibit Smad3/4 and TGF- β signalling by directly binding and sequestering Smad3, or by interfering with

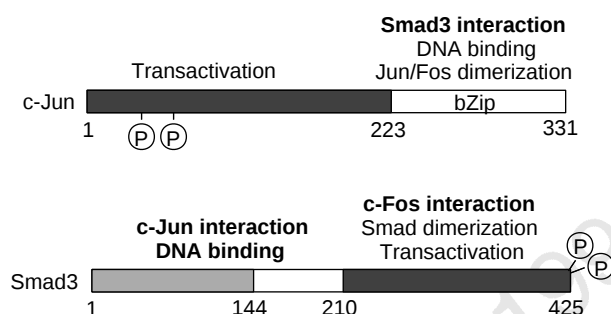


Figure 5 Domains of c-Jun and Smad3. The functions shown in bold have been assigned on the basis of our present results.

the binding of Smad3 to the DNA.

Overexpression of the N–L segment of Smad3 also inhibited TGF- β -induced transcription from the TRE promoter in Mv1Lu cells (Fig. 4b). Although this inhibition may be due to ability of the N–L domain to interact with the C-domain of Smad3 and thus prevent Smad3 association with Smad4 (ref. 18), it may also be due to its interaction with c-Jun and/or the TRE, thus creating an inactive complex. Accordingly, Smad3N–L, but not Smad1N–L, inhibited the transcriptional activity of c-Jun and c-Jun–c-Fos (Fig. 4c) and interfered with the interaction between c-Jun and c-Fos in mammalian two-hybrid assays (Fig. 4d).

Our results suggest a model for how Smad3 and Smad4 mediate TGF- β -induced transcription at AP1-binding sites through physical and functional interactions with the TRE and with c-Jun/c-Fos. Thus, Smad3 and Smad4 associate with each other and interact directly with the Smad-binding sequence at the AP1-binding site in a conceptually similar manner to c-Jun–c-Fos heteromers^{9,24}. The resulting Smad3/4-mediated transcription is highly TGF- β -inducible and does not require c-Jun or c-Fos, as shown in F9 cells (Fig. 1b). In the presence of c-Jun and c-Fos, the heteromeric Smad complex interacts in a ligand-dependent fashion with the AP-1 complex, primarily through interaction of Smad3 with c-Jun, but probably also stabilized by the affinity of Smad3 for the DNA and for c-Fos and Smad4. The resulting synergy of Smad3–Smad4 and AP-1 then confers strong, highly TGF- β -inducible transcription from the AP1-binding site. Our results also indicate that the functional organization of Smad3 resembles that of the bZIP transcription factors (Fig. 5), which have a basic region that mediates DNA binding closely preceding the leucine-zipper dimerization domain, and a transactivation domain located at the other end of the protein^{9,24}. Whereas bZIP transcription factors function by dimerization, further functional diversity can be achieved from combinatorial pairwise interactions between Smad proteins and AP-1 transcription factors.

Finally, mitogenic activation of tyrosine kinase receptors, stress and ultraviolet irradiation all activate MAP kinase cascades and JNK kinase, leading to transcriptional activation of the AP-1 complex²⁴. Our results show that these signals converge with TGF- β -induced Smad signalling at the AP-1 promoter site. The phosphorylation of Smad1 and Smad2 (refs 26, 27) in response to mitogenic stimulation of MAP kinase raises the possibility that such phosphorylation of Smad3 may regulate its interaction with the AP1-binding site and the AP-1 complex. □

Methods

Expression plasmids and reporter constructs. N-terminal Flag-tagged Smad1–3, haemagglutinin (HA)-tagged or VP16-fused c-Jun and c-Fos, C-terminal Flag-tagged Smad1NL (amino acids 1–245), and truncated c-Jun (amino acids 1–223) and c-Jun (amino acids 235–331) were generated by polymerase chain reaction (PCR)-based subcloning into plasmid pRK5 (ref. 28). Gal4(1–147)-fused Smad3 was made by inserting the Smad3 complemen-

tary DNA into pSG424 (ref. 29). Details can be provided on request. pGal4–Smad1 and pGal4–Smad4 were provided by J. Massagué, and pRSV–c-Jun, pRSV–c-Fos and the pRSV vector plasmids were provided by R. Tjian and K. Yamamoto. The C-tagged pRK5–Smad4F (ref. 3), pRK5–Smad3NL (ref. 4) and constitutively active cytoplasmic TGF- β -receptor (T β R)II–T β RI have been described³⁰. The -73Col-Luc reporter plasmid contains the luciferase gene under control of a truncated collagenase I promoter with a single AP-1 consensus sequence¹³, whereas TRE-Luc contains four tandem copies of the AP-1 consensus sequence in front of the luciferase gene¹². pFR-Luc has five copies of a Gal4-binding element, followed by the luciferase gene (Stratagene).

Transient transfections and functional assays. Transient transfections, TGF- β treatment, Gal4 transactivation, mammalian two-hybrid assays and luciferase assays were done as described¹⁷. For each transfection, 0.5 μ g of each expression plasmid, 0.5 μ g luciferase reporter plasmid, and 0.25 μ g β -galactosidase plasmid were used. For the Gal4 transactivation assay and mammalian two-hybrid assays, 50 ng Gal4(1–147) fusion plasmids were used. The total plasmid concentration was kept constant, and, when needed, vector DNA was added.

GST-fusion proteins and *in vitro* protein-binding assays. Plasmids pGEX–Smad3 and pGEX–Smad4 have been described³ and all other GST–Smad fusion were made by subcloning the Smad cDNAs from corresponding pRK5–Smad plasmids⁴ into pGEX (Pharmacia). Equal amounts of GST or GST–Smad fusion protein bound to glutathione–Sephadex beads were incubated with ³⁵S-labelled, *in vitro*-translated proteins (TNT translation kit, Promega) with similar specific radioactivity. associated ³⁵S-labelled proteins were detected by SDS–PAGE and autoradiography.

DNA affinity purification of associated proteins, immunoprecipitation, western blotting. For affinity purification of proteins bound to the AP-1-binding oligonucleotide, 200 ng of a biotinylated AP-1 probe were immobilized to streptavidin-conjugated magnetic beads (Promega). The beads were then incubated with lysates of transiently transfected COS-1 cells. After extensive washing, the AP-1 probe and associated proteins, immobilized on magnetic beads, were removed using a magnet and resuspended in SDS sample buffer. Flag-tagged Smad3 and Myc-tagged Smad4 were detected by immunoblotting using anti-Flag M2 antibody (Kodak, IBI) or anti-Myc 9E10 antibody. c-Jun was detected using a mixture of two anti-c-Jun antibodies, c-Jun(D) and c-Jun(N), and c-Fos was detected by using c-Fos(4)–horseradish peroxidase (HRP) (Santa Cruz Biotech). Immunoprecipitations were performed as described³⁰. Sequentially immunoprecipitated complexes were visualized by autoradiography and Smad3-associated endogenous c-Jun was visualized by immunoblotting as described above.

Electrophoretic mobility-shift assays. Mobility-shift assays were done using a Promega gel-shift assay kit. The reactions contained 0.4–0.5 μ g glutathione–Sephadex-purified GST fusion protein or purified c-Jun (Promega). Excess unlabelled competitor oligonucleotide (Promega) was added to the reaction mixture before preincubation, when required. For gel mobility shift and supershift assays using cell nuclear extracts, 1 μ l extract containing about 1 μ g μ l⁻¹ protein, prepared as described²², was used. For supershift analyses, 1 μ l antibody, HA(12A5) against HA–c-Jun (Babco) and antibodies against c-Fos(K-25) (Santa Cruz Biotech), affinity-purified Smad3 (ref. 11) and Smad4 (ref. 22), were incubated at 4 °C for 90 min after adding ³²P-labelled probes.

DNA footprinting. The DNA probe for DNaseI footprinting containing two tandem AP-1-binding sequences (sequence shown in Fig. 3d) were excised from pBSK2XAP-1, a derivative of pBS2SK⁺ (Stratagene), and 5'-³²P-labelled at only one end. Footprinting was carried out using the core footprinting system (Promega).

Received 10 June; accepted 27 July 1998.

- Heldin, C. H., Miyazono, K. & ten Dijke, P. TGF- β signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471 (1997).
- Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massagué, J. Partnership between DPC4 and SMAD proteins in TGF- β signalling pathways. *Nature* **383**, 832–836 (1996).
- Zhang, Y., Feng, X.-H., Wu, R.-Y. & Derynck, R. Receptor-associated Mad homologues synergize as effectors of the TGF- β response. *Nature* **383**, 168–172 (1996).
- Zhang, Y., Musci, T. & Derynck, R. The tumor suppressor Smad4/DPC4 as a central mediator of Smad function. *Curr. Biol.* **7**, 270–276 (1997).
- de Groot, R. P. & Kruijer, W. Transcriptional activation by TGF- β mediated by the DYAD symmetry element (DSE) and the TPA responsive element (TRE). *Biochem. Biophys. Res. Commun.* **168**, 1074–1081 (1990).
- Keeton, M. R., Curriden, S. A., van Zonneveld, A. & Loskutoff, D. J. Identification of regulatory sequences in the type 1 plasminogen activator inhibitor gene responsive to transforming growth factor β . *J. Biol. Chem.* **266**, 23048–23052 (1991).
- Jin, G. & Howe, P. H. Regulation of clusterin gene expression by transforming growth factor β . *J. Biol. Chem.* **272**, 26620–26626 (1997).
- Felts, S. J., Stoflet, E. S., Eggers, C. T. & Getz, M. J. Tissue factor gene transcription in serum-stimulated fibroblasts is mediated by recruitment of c-Fos into specific AP-1 DNA-binding complexes. *Biochemistry* **34**, 12355–12362 (1995).
- Angel, P. & Karin, M. The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. *Biochim. Biophys. Acta* **1072**, 129–157 (1991).
- Wrana, J. L. et al. TGF- β signals through a heteromeric protein kinase receptor complex. *Cell* **71**, 1003–1014 (1992).
- Nakao, A. et al. TGF- β receptor-mediated signaling through Smad2, Smad3 and Smad4. *EMBO J.* **16**, 5353–5362 (1997).
- Renshaw, M. W., McWhirter, J. R. & Wang, J. Y. The human leukemia oncogene bcr-abl abrogates the anchorage requirement but not the growth factor requirement for proliferation. *Mol. Cell. Biol.* **15**, 1286–1293 (1995).
- Angel, P. et al. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell* **49**, 729–739 (1987).
- Wu, R.-Y., Zhang, Y., Feng, X.-H. & Derynck, R. Heteromeric and homomeric interactions correlate with signaling activity and functional cooperativity of Smad3 and Smad4/DPC4. *Mol. Cell. Biol.* **17**, 221–228 (1997).
- Chiu, R. et al. The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. *Cell* **54**, 541–552 (1988).
- Yang-Yen, H. F., Chiu, R. & Karin, M. Elevation of AP-1 activity during F9 cell differentiation is due to increased c-Jun transcription. *New Biol.* **2**, 351–361 (1990).
- Feng, X.-H., Zhang, Y., Wu, R.-Y. & Derynck, R. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for Smad3 in TGF- β -induced transcriptional activation. *Genes Dev.* **12**, 2153–2163 (1998).
- Hata, A., Lo, R. S., Wotton, D., Lagna, G. & Massagué, J. Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4. *Nature* **388**, 82–87 (1997).
- Liu, F. et al. A human Mad protein acting as a BMP-regulated transcriptional activator. *Nature* **381**, 620–623 (1996).
- Feng, X.-H., Filvaroff, E. H. & Derynck, R. TGF- β -induced down-regulation of cyclin A expression requires a functional TGF- β receptor complex. Characterization of chimeric and truncated type I and type II receptors. *J. Biol. Chem.* **270**, 24237–24245 (1995).
- Yoshizumi, M. et al. Down-regulation of the cyclin A promoter by transforming growth factor- β 1 is associated with a reduction in phosphorylated activating transcription factor-1 and cyclic AMP-responsive element-binding protein. *J. Biol. Chem.* **272**, 22259–22264 (1997).
- Yingling, J. M. et al. Tumor suppressor Smad4 is a transforming growth factor beta-inducible DNA binding protein. *Mol. Cell. Biol.* **17**, 7019–7028 (1997).
- Zawel, L. et al. Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol. Cell* **1**, 611–176 (1998).
- Karin, M. The regulation of AP-1 activity by mitogen-activated protein kinases. *J. Biol. Chem.* **270**, 16483–16486 (1995).
- Lloyd, A., Yancheva, N. & Wasylyk, B. Transformation suppressor activity of a Jun transcription factor lacking its activation domain. *Nature* **352**, 635–638 (1991).
- Kretschmar, M., Doody, J. & Massagué, J. Opposing BMP and EGF signalling pathways converge on the TGF- β family mediator Smad1. *Nature* **389**, 618–622 (1997).
- de Caestecker, M. P. et al. Smad2 transduces common signals from serine-threonine and tyrosine kinases. *Genes Dev.* **12**, 1587–1592 (1998).
- Graycar, J. L. et al. Human transforming growth factor- β 3: recombinant expression, purification, and biological activities in comparison with transforming growth factor- β 1 and - β 2. *Mol. Endocrinol.* **3**, 1977–1986 (1989).
- Sadowski, I. & Ptashne, M. A vector for expressing GAL4(1–147) fusions in mammalian cells. *Nucleic Acids Res.* **17**, 7539 (1989).
- Feng, X.-H. & Derynck, R. Ligand-independent activation of TGF- β signaling pathways by heteromeric cytoplasmic domains of TGF- β receptors. *J. Biol. Chem.* **271**, 13123–13129 (1996).

Acknowledgements. This research was supported by the NIH (R.D.) and postdoctoral fellowships from the American Lung Association (to Y.Z.) and the American Cancer Society (to X.-H.F.). We thank J. Massagué, J. Wrana and L. Attisano, M. Karin, D. Loskutoff, J. Wang, K. Yamamoto, P. Kushner, X.-F. Wang, P. ten Dijke, M. Green, M. Ptashne and R. Tjian for valuable reagents, and M. Cronin and L. Choy for reading of the manuscript.

Correspondence and requests for materials should be addressed to R.D. (e-mail: derynck@itsa.ucsf.edu).

Biochemical reagents on tap

Research supplies are in need of restocking and the necessary items may include factor Xa, D-amino acids, trichostatin A, scyllo-inositol, or selected antibodies and labelling reagents.

Proteins and peptides

From Cambio

New additions to the company's range of proteins and peptides

New additions include the HIV gag, nef, env, p24 and gp120 proteins, recombinant peptides from the herpes NS3 and NS4 proteins, the hepatitis C virus-core protein and several new FAS antigen reagents. All the new antibodies are available unmodified, or with a range of biotin, FITC or rhodamine labels. The company offers hundreds of antibodies against many pathogens and proteins, mycobacteria, *Chlamydia*, TIMP, collagenase, acholeplasma, CFTR and huntingtin, to name a few. All antibodies have a stated purity of at least 95 per cent.

Reader Enquiry No. 100

Trichostatin A

From Wako BioProducts

This is a specific inhibitor for histone deacetylation — now available in a 1-mg vial

This antimitotic drug is known to arrest mitosis of animal cells at G1 and G2, says Wako. At relatively low concentrations, the drug inhibits the deacetylation of histone and is useful for verifying the regulation of chromatin structure relative to cellular dynamics. Also offered by Wako are synthetic retinoids for biochemistry research (Am80, Am580, Re80, Ch55 and LE540). These synthetic retinoids should find use in the study of the signal transduction system that uses retinoid receptors. Retinoids and retinoic acid are also involved in cell differentiation, proliferation and embryonic development in vertebrates.

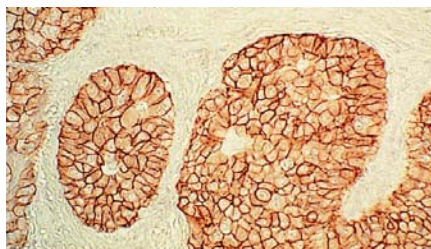
Reader Enquiry No. 101

mAb to c-erbB-2 external domain

From Zymed

TAB250 mouse monoclonal antibody to the extracellular domain of the c-erbB-2 protein

The extracellular domain is also recognized by Herceptin, a humanized anti-c-erbB-2 mAb in clinical trials for breast cancer therapy. TAB250 produces vivid plasma membrane-associated staining with formalin-fixed, paraffin-embedded or frozen tissue sections, without the need for tissue pretreatment with heat or enzymes. TAB250 has proven useful in clinical research studies and is now available in a prediluted format for manual and automated immunostaining. The c-erbB-2 oncogene is amplified and over-expressed in as many as 30 per cent of



Zymed's mouse c-erbB-2 mAb (external domain).

breast cancers. Several studies correlate c-erbB-2 over-expression with tumour aggressiveness, poor response to endocrine-related therapy and a poor prognosis.

Reader Enquiry No. 102

Splitmix

From Biomedica

A non-enzymatic cell disruption reagent available in the UK through Autogen Bioclear

This reagent can be used for the disruption of cell monolayers as an alternative to trypsin. Splitmix is of plant origin and its gentle non-enzymatic cell detachment minimizes cell damage, which is in contrast to trypsinization, says Biomedica. It is prepared in PBS solution without calcium and magnesium, and is available in two concentrations (8 g l^{-1} and 14 g l^{-1}).

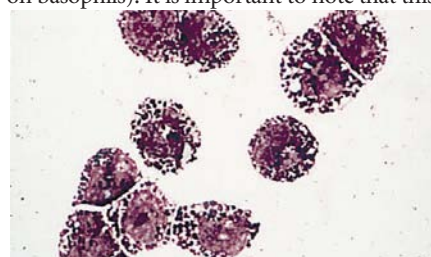
Reader Enquiry No. 103

StemSep basophils enrichment

From StemCell

An immunomagnetic system for the isolation of human basophils from peripheral blood

The procedure is said to be rapid and simple, with an average stated recovery of 87 per cent IgE^+ cells. The purified basophils are not labelled with antibody and are well suited for the study of basophil activation. This negative selection system captures unwanted cells, and the desired cells flow through the magnetic column. For the enrichment of basophils from human peripheral blood, a depletion cocktail of antibodies to haematopoietic lineage markers is provided (these are not found on basophils). It is important to note that this



StemCell enriches basophils with magnetism.

cocktail will not remove mast cells. Both basophils and mast cells are enriched; however, adherent, damaged, or dead cells are automatically removed. In addition, column capacities range from 2×10^7 to 1.5×10^9 nucleated cells.

Reader Enquiry No. 104

Factor Xa

From Enzyme Research Laboratories

A common molecule to both the intrinsic and extrinsic clotting systems, which plays an important role in blood coagulation

Purified human and bovine factor Xa is now readily available in quantities and specifications to suit research requirements. In addition, the company provides a number of coagulation reagents to scientists involved in thrombosis and haemostasis research, including purified coagulation factors, antibodies to coagulation factors and purified antibodies for ELISAs.

Reader Enquiry No. 105

D-phenylalanine and D-tyrosine

From NSC Technologies

D-amino acid transaminase is used to produce these enantiomers on a multi-tonne scale

Drawing on the L-phenylalanine fermentation process, the new process was evolved in response to increasing demand from the pharmaceutical industry. According to the manufacturer, D-phenylalanine and D-tyrosine are currently used as intermediates in a wide range of drugs.

Reader Enquiry No. 106

Chemokine reagents

From Serotec

A wide range of antibodies and recombinant proteins: human, rodent and ovine

Serotec offers human, mouse, rat and sheep chemokines, chemokine receptor monoclonal antibodies, polyclonal antibodies and recombinant proteins. Applications for these reagents include flow cytometry, immunocytochemistry, ELISA, western blotting, immunoprecipitation and functional assays.

Reader Enquiry No. 107

What's in a label

Fluorochrome conjugated antibodies

From Miltenyi Biotec

Various FITC- or PE-conjugated antibodies

Developed for the convenient evaluation of MACS magnetic cell separations, these anti-

bodies can also be used for staining unseparated cells. According to the manufacturer, they are well suited for flow cytometric or fluorescent microscopic analysis, even of large cell numbers. The ready-to-use fluorochrome conjugated antibodies are optimized for 100 tests of up to 10^7 cells. The specificity of the antibodies is said to provide excellent results for the identification and enumeration of individual cell types.

Reader Enquiry No. 108

Phycobilisome-labelled reagents

From KPL

A new line of streptavidin reagents labelled with PBXL-1 and PBXL-3

These reagents exhibit a large Stokes shift with red fluorescence, and they are said to eliminate the need for substrates by offering direct fluorescent labelling. The dye complex consists of B-phycoerythrin, R-phycoerythrin and allophycocyanin, which can be used with ELISA, western blotting, DNA arrays, fluorescence microscopy and flow cytometry.

Reader Enquiry No. 109

Immunosubstrates

From Bio FX

New non-hazardous immunosubstrate formulations for TMB, ABTS and BCIP/NBT

These substrate formulations are for use with peroxidase (TMB, ABTS) and alkaline phosphatase (BCIP) enzyme conjugates and have been developed to provide improvements in sensitivity, as well as the stability of one-component products. The system is said to have ultra low background absorbance combined with the increased sensitivity, improving the signal-to-noise ratio. The production protocols use buffers that are said to improve lot-to-lot consistency. Moreover, stability and functionality is accomplished without the use of polar aprotic solvents, thus reduces the biohazard risks normally associated with immunosubstrates. These substrates come ready-to-use and packaged to specific needs.

Reader Enquiry No. 110

PtdIns (3,4,5)-P₃

From Alexis

A key second messenger in cell regulation

This new synthetic phosphatidylinositol 3,4,5-triphosphate is believed to be identical to naturally occurring compounds. It is not an analogue containing only saturated fatty acid residues. Moreover, this material displays significantly greater biological activity than such saturated analogues, says Alexis.

Reader Enquiry No. 111

Receptors and reagents

From Research Biochemicals and Semat

A range of receptors, receptor antagonists, receptor antibodies and inhibitors

Products introduced during the last year include a standard, selective D₁ antagonist SCH 23390 hydrochloride; a selective, competitive and orally active GABA_B SCH50911; and hydroxylamine hydrochloride, a cell permeable catalase-dependent NO precursor (the use of which as an NO source mimics more closely the events following an increase in intracellular NO production).

Reader Enquiry No. 112

Signal transduction lipids

From Sigma

Inositol phosphates (IP) and other bioactive lipids for signal transduction research

A broad selection of lipids for use in signal transduction research is available, including over 30 myo-inositol derivatives, such as IP₄, inositol 1,4,5-phosphate, scyllo-inositol (DTLET), and other myo-inositol phosphate isomers. These lipids meet the company's quality control and purity requirements, and are backed with comprehensive technical support. Package for laboratory and manufacturing use are available.

Reader Enquiry No. 113

These notes are compiled by Brendan Horton from information provided by the manufacturers.

For more details, fill in the reader service card bound inside the journal.

PIERCE

B-PER™ Bacterial Protein Extraction Reagent

Streamlines Protein Extraction From Bacteria

Features/Benefits:

- Eliminates the need for mechanical separation via equipment that isn't always available
- Compatible with affinity purifications
- Purifies inclusion bodies to near homogenous levels and extracts all possible soluble proteins
- Easy-to-use, one-reagent formula eliminates external protein contamination by the lysis reagent
- Can also be used with baculovirus-expressed proteins

Tel: 800-874-3723 or 815-968-0747 • Fax: 815-968-7316
CS@piercenet.com • Internet: www.piercenet.com

READER ENQUIRY NO. 44

Lysyl Endopeptidase (Achromobacter Protease 1)

*Specifically cleaves all Lys-X bonds including Lys-Pro.
Highly Stable, Water Soluble*



Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 31

ANTIBODIES

For Research & Routine Histology

*Information - Angiogenesis - Tumors
Leukocytes - Endothelial Cells
Pathological Markers*

Custom Monoclonals and Polyclonals

BMA Biomedicals Ltd.
(800) 634-3183 (US)
Int: +41 61 811 62 22
+41 61 811 60 06 (fax)
info@bma.ch

READER ENQUIRY NO. 28

serotec
...better products, better service, better quality...

- Antibodies to Human CD3
- Antibodies to Mouse CD3
- Antibodies to Rat CD3
- Antibodies to Guinea CD3
- Antibodies to Bovine CD3
- Veterinary Research Antibodies
- Antibodies to Hormones
- Antibodies to Microbial Antigens
- Apoptosis Reagents
- Adhesion Molecule Products
- Cytokine & Growth Factor Reagents
- Cytoskeletal Component Reagents
- Flow Cytometry Reagents
- Hollow Fiber Bioreactors
- Cell Signaling Reagents
- Neurological Reagents
- Angiogenesis Reagents
- Reproductive Biology Reagents (Fertilizing)
- TDR Reagents
- Tissue Culture Products

100% RESEARCH USE ONLY

Services: 22 Bedford, Station Approach, Kington, Oxford OX5 1B, UK
Tel: +44 (0)1865 812700 Fax: +44 (0)1865 812701
Technical Enquiries Tel: +44 (0)1865 812711
Order Routes Tel: +44 (0)1865 812722
www.serotec.co.uk

READER ENQUIRY NO. 23

PACEMAKER®

Quality assured
BS:EN ISO9002
accredited

Custom peptides, conjugates and antibodies

Unrivalled experience and expertise with assured quality

*For more information on AFFINITI's Peptide and Protein Research Services,
please contact:*

AFFINITI Research Products Ltd., Mamhead Castle, Exeter, EX6 8HD, UK
Tel. +44/0 1626 891010; Fax: +44/0 1626 891090; URL: <http://www.affiniti-res.com>

READER ENQUIRY NO. 32